I. S. E. G.
Inf. Biblioteca
286-6. 44981
QA76.87.C67 1996
X-96-055652-3
RESERVADO
I. S. E. G. BIBLIOTECA

UNIVERSIDADE TÉCNICA DE LISBOA

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

Mestrado em: Matemática Aplicada à Economia e à Gestão

Redes Neuronais no Âmbito da Realidade Virtual

(A Resolução do Problema da Obtenção da Estrutura Tridimensional dos Objectos)

Paulo Jorge Duarte Correia

Orientação: professor doutor Ilidio Antunes

Júri:
Presidente -professor doutor Ilidio Antunes
Vogais -professor doutor Armando Alves de Oliveira
-dr. Raúl Massano Brás

Outubro de 1996

Errata

Pag. 93: a expressão (4.9) deve ser substituída por:

$$\Leftrightarrow \sum_{i=1}^{N_l}\left(1-\sum_{k=1}^{N_r}P_{ik}\right)^2=\sum_{i=1}^{N_l}(1)+\sum_{i=1}^{N_l}\left(\sum_{k=1}^{N_r}\sum_{l=1}^{N_r}P_{ik}P_{il}\right)-\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}2P_{ik} \Leftrightarrow$$

Pag. 93: a expressão (4.13) deve ser substituída por:

$$\Leftrightarrow \sum_{k=1}^{N_r}\left(1-\sum_{i=1}^{N_l}P_{ik}\right)^2=\sum_{k=1}^{N_r}(1)+\sum_{k=1}^{N_r}\left(\sum_{i=1}^{N_l}\sum_{j=1}^{N_l}P_{ik}P_{kj}\right)-\sum_{k=1}^{N_r}\sum_{i=1}^{N_l}2P_{ik} \Leftrightarrow$$

Pag. 93: a expressão (4.14) deve ser substituída por:

$$\Leftrightarrow \sum_{k=1}^{N_r}\left(1-\sum_{i=1}^{N_l}P_{ik}\right)^2=N_r+\sum_{k=1}^{N_r}\sum_{i=1}^{N_l}\sum_{j=1}^{N_l}P_{ik}P_{kj}-\sum_{k=1}^{N_r}\sum_{i=1}^{N_l}2P_{ik}$$

Pag. 93: onde se lê "Substituindo, em (4.14), $P_{ik}P_{il}$ ..." deve-se ler "Substituindo, em (4.14), $P_{ik}P_{kj}$ ..."

Pag. 94: a expressão (4.20) deve ser substituída por:

$$C_{ikjl}=\frac{2}{\left[1+e^{\lambda(X-\theta)}\right]}-1$$

Pag. 99, ante-penúltima linha: deve-se ler $M(t-1)=\{\hat{P}_i(t-1): i=1,\ldots,N\}$

Pag. 99, penúltima linha: deve-se ler $\hat{P}_i(t-1)$

Pag. 100: a expressão (4.27) deve ser substituída por:

$$F=\sum_{i,j:i\neq j}\eta_{ij}\left(D\left[\hat{P}_i(t-1),\hat{P}_j(t-1)\right]-D\left[\hat{P}_i(t),\hat{P}_j(t)\right]\right)^2$$

Pag. 102: a expressão (4.27) deve ser substituída por:

$$M(t-1)=d_{ij}(t-1)=D\left[\hat{P}_i(t-1),\hat{P}_j(t-1)\right]; i=1,\ldots,N; j=1,\ldots,N; i\neq j$$

Pag. 102: a expressão (4.28) deve ser substituída por:

$$w_{ij}(t)=\Delta\hat{Z}_{ij}(t)=\left|\hat{Z}_i(t)-\hat{Z}_{ij}(t)\right|=\sqrt{d_{ij}^2(t-1)-\left[X_i(t)-X_j(t)\right]^2-\left[Y_i(t)-Y_j(t)\right]^2}$$

1996

UNIVERSIDADE TÉCNICA DE LISBOA

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

Redes Neuronais no Âmbito da Realidade Virtual

(A Resolução do Problema da Obtenção da Estrutura Tridimensional dos Objectos)

Orientação: professor doutor Ilidio Antunes

Júri:

Presidente	-professor doutor Ilidio Antunes
Vogais	-professor doutor Armando Alves de Oliveira
	-dr. Raúl Massano Brás

Glossário

Características: certas propriedades presentes nas imagens que se destacam e permitem a detecção de objectos nelas presentes.

Computador: sistema electrónico capaz de receber instruções para aceitar, processar, armazenar e apresentar dados e informação.

Disparidades: correspondem à diferença entre as coordenadas de um ponto de uma imagem e as coordenadas do ponto correspondente da outra imagem.

Intensidade da luminosidade: as imagens monocromáticas são compostas por pontos cuja luminosidade varia numa escala de tons de cinzento. A intensidade da sua luminosidade corresponderá a um valor particular nessa escala.

Imersão: processo pelo qual, mediante a utilização de equipamento apropriado, o utilizador tem a sensação de estar fisicamente inserido num ambiente virtual. Para alcançar tal efeito, o equipamento procura iludir os sentidos humanos, substituindo a percepção do mundo real pelas sensações provenientes do ambiente virtual.

Interacção: processo pelo qual o utilizador consegue desenvolver acções que irão agir sobre a informação que lhe é apresentada, permitindo-lhe também receber *feedback* resultante dessas acções.

Interface: componente de um sistema computacional que permite ao utilizador comunicar com a unidade de processamento central.

Know-how: capacidade para a realização de determinada tarefa.

Neurónios artificiais: modelo matemático inspirado no funcionamento dos neurónios biológicos destinado ao processamento de informação. Tal como os seus homólogos

biológicos, a sua formologia permite-lhes receber e transmitir impulsos de e para outros neurónios.

Neurónios biológicos: células que integram o sistema nervoso, compostas pelo corpo celular, dendrites e axónio, que são responsáveis pela transmissão de impulsos nervosos através de conexões que estabelecem entre si.

Nodos: parte central de um neurónio artificial, na qual é efectuada uma computação dos impulsos recebidos de outros neurónios.

Oclusão: fenómeno que ocorre quando um ponto está visível numa imagem e desaparece na imagem seguinte, ou porque está tapado por outros objectos, ou porque saiu do campo de visão da câmara.

Projecção: processo pelo qual os objectos do mundo real são tranferidos para o plano bidimensional da imagem captada pela câmara.

Realidade ampliada: tecnologia que, através da utilização de óculos transparentes, permite ao seu utilizador sobrepôr informação adicional à visão que tem do ambiente que o rodeia.

Redes de comunicação: Consistem numa ligação entre vários locais através de um meio que permita aos seus utilizadores trocarem dados e informação.

Redes neuronais artificiais: são compostas por um conjunto de neurónios artificiais entre os quais se estabelecem ligações para a transmissão de impulsos. A arquitectura da rede permite-lhe tomar decisões para a resolução dos problemas que lhe são colocados.

Rígidez: característica dos objectos que mantêm a sua forma inalterada, ou seja, as distâncias entre os vários pontos que os compõem mantêm-se constantes.

RV: realidade virtual. Conjunto de tecnologias que têm por objectivo proporcionar ao seu utilizador novas formas de percepcionar e interagir com a informação.

Sistemas de informação: conjunto de componentes, formando um todo, que interagem por forma a permitir a troca de dados e informação entre os utilizadores.

Telepresença: conjunto de tecnologias que permitem ao utilizador interagir à distância com um meio ambiente, real ou virtual, no qual não está fisicamente presente.

TI: tecnologia(s) de informação. Conjunto de items e capacidades utilizados para a criação, armazenamento e dispersão de informação.

UE: unidades económicas. São todos os entes sociais organizados e reconhecidos por lei, com fins lucrativos ou não, que fazem parte integrante de uma sociedade.

Visão estereoscópica: consiste num processo de visão que apresenta uma geometria particular que permite obter uma visualização a três dimensões da cena observada.

Resumo

O objectivo deste trabalho é propôr um modelo para a obtenção da estrutura tridimensional de uma cena baseado na utilização de redes neuronais que possa ser aplicado no desenvolvimento de sistemas de Telepresença. O problema que se coloca é a obtenção da estrutura tridimensional de uma cena a partir da análise das suas projecções numa sequência de imagens bidimensionais. Como tal , trata-se de um problema de Visão por Computador, que envolve três fases: detecção de características presentes nas imagens, estabelecimento de correspondências entre os pontos característicos e obtenção da estrutura tridimensional.

Neste trabalho são estudadas e analisadas Tecnologias de Informação consideradas inovadoras: a Realidade Virtual e as Redes Neuronais. No caso da Realidade Virtual, são descritas as suas características e potencialidades e é feito um levantamento de alguns campos de aplicação. De seguida, é efectuado um paralelismo com as modernas Tecnologias de Informação e Sistemas de Informação e são especificados alguns dos efeitos que poderá exercer sobre as Unidades Económicas.

É efectuado um levantamento de alguma pesquisa levada a cabo no domínio da Visão por Computador como forma de introdução a esta problemática.

De seguida, são apresentadas as redes neuronais para a resolução do problema. Para o estabelecimento de correspondências é utilizada uma rede de Hopfield. Define-se uma função de custo que incorpora as restrições do problema, a qual irá ser minimizada pela rede. Cada neurónio representa uma possível correspondência entre um ponto de uma imagem e um ponto da imagem seguinte.

Por fim, apresenta-se a rede neuronal para a obtenção da estrutura tridimensional a partir das correspondências. É seguido o princípio da máxima rígidez, à semelhança do que acontece com o sistema visual humano. Existe um modelo interno da estrutura tridimensional, que será actualizado gradualmente, permitindo a existência de alguns desvios de rígidez.

Palavras-chave: Redes Neuronais, Realidade Virtual, Estrutura Tridimensional, Visão por Computador, Tecnologias de Informação

Abstract

The purpose of this work is to propose a model for the recovery of the tridimensional structure of a scene based on neural networks. This model may be used to develop Telepresence systems. The problem that we face is to determine the 3-D structure of a scene from an analysis of its projections on a sequence of 2-D images. This is a Computer Vision problem that involves three phases: features detection, matching of the feature points, and computation of the 3-D structure.

In this work we study and analyse Information Technologies which are considered innovative: Virtual Reality and Neural Networks. In the case of Virtual Reality, we describe some of its characteristics and potentials and we present some of its possible applications. Next, we compare Virtual Reality to the modern Information Technologies and Information Systems and we suggest some of the possible effects of its use by Economic Units.

We refer some of the work done in the field of Computer Vision so that we can show some of the problems that researchers in this field have to face.

Next, we state the neural networks that will be used to solve the problem. For the matching phase we suggest the use of an Hopfield neural network. We state a cost function that will represent the constraints on the solution, which will be minimised by the network. Each neurone represents a possible match between a point of an image and a point in the next image.

Then, we present the neural network that will be used to compute the 3-D structure from the correspondences established. That computation is constrained by the maximal rigidity principle, a process similar to the way we see. There is an internal model of the 3-D structure that will be updated at each instant, allowing for some deviations from rigidity.

Index Terms: Neural Networks, Virtual Reality, 3-D Structure, Computer Vision, Information Technologies

Lista de Figuras e Tabelas

Figuras

Tabelas

Agradecimentos

Em primeiro lugar, gostaria de agradecer à minha família e amigos pela compreensão demonstrada e pelo constante apoio que me deram.

Em segundo lugar, quero agradecer ao orientador deste trabalho, o professor doutor Ilidio Antunes, pela sua disponibilidade e pelas observações e sugestões que me deu.

Por último, gostaria de agradecer à Direcção da Extensão de Portimão do ISMAG pelas facilidades que me foram concedidas a nível de horário, permitindo-me ter tempo disponível para a elaboração desta Dissertação.

ÍNDICE

1. Introdução

Nesta dissertação abordamos o tema das redes neuronais. Trata-se de um ramo do saber que já existe há algumas décadas, tendo conhecido sucessivas fases de avanços e retrocessos. Actualmente, vive-se uma fase em que os investigadores estão a redescobrir as suas potencialidades para aplicação à resolução de novos problemas. Um dos campos de aplicação mais recentes consiste na análise de imagens e, em particular, à inferência acerca da tridimensionalidade dessas imagens. Uma vez que a tridimensionalidade desempenha um papel fulcral na tecnologia da Realidade Virtual (RV), decidimos aplicar as redes neuronais à obtenção de um modelo que possa ser utilizado nesse domínio.

Pensamos que a RV apresenta características que certamente irão provocar alterações profundas a nível das Tecnologias de Informação (TI) e dos Sistemas de Informação (SI). A sua forma peculiar de apresentar a informação irá certamente contribuir para que esta seja assimilada pelos seus utilizadores de uma forma muito mais eficiente e intuitiva. A sua utilização poderá constituir um factor crítico de sucesso, comparável ao que acontece hoje em dia com os computadores.

De facto, uma vez que vivemos na chamada Era da Informação, qualquer factor que permita utilizar a informação de uma forma mais eficiente permitirá a obtenção de vantagens competitivas. As Unidades Económicas (UE) poderão obter grandes proveitos com a sua utilização. A partir do momento em que a tecnologia da RV esteja suficientemente desenvolvida ao ponto de estar disponível para um consumo de massas, como se passa hoje em dia com os computadores, o panorama das TI irá mudar radicalmente.

Uma vez que este trabalho pretende abordar as TI (em termos gerais), a associação entre RV e redes neuronais parece-nos particularmente interessante, visto que estas últimas também podem ser consideradas uma tecnologia destinada ao tratamento de informação. De facto, as redes neuronais são utilizadas para a resolução de problemas através de uma análise de informação que lhes é fornecida. Essa informação vai ser processada e analisada no seu interior, permitindo-lhes obter a informação final desejada pelo utilizador do algoritmo.

O objectivo deste trabalho consiste em definir uma aplicação das redes neuronais ao domínio da RV. Em particular, pretendemos sugerir a sua utilização para a formalização de um modelo que permita obter uma representação tridimensional de uma cena a partir da análise de imagens bidimensionais captadas dessa mesma cena. Tal modelo poderá ser particularmente útil para a concepção de sistemas de Telepresença. Ou seja, sistemas que, através da utilização de uma câmara, captem imagens de um certo ambiente, à distância, e as enviem para o utilizador para que ele, do local onde se encontra, possa interagir, em termos reais ou em termos virtuais, com elementos presentes nesse mesmo ambiente. Para tal fim, o modelo deverá computar uma representação tridimensional virtual da cena que estiver a ser visionada. Isso será levado a cabo através da análise das imagens obtidas pela câmara. O problema que se coloca é que, através do processo de captação das imagens, toda a informação relativa à profundidade da cena será perdida. De facto, a única informação disponível consiste nas projecções dos objectos presentes na cena no plano bidimensional das imagens obtidas. Portanto, a posição e estrutura dos objectos apenas podem ser identificadas pelas suas coordenadas bidimensionais. O problema consiste na recuperação da terceira coordenada, aquela que dá informação respeitante à profundidade dos objectos.

A principal tarefa do modelo consistirá então na obtenção das coordenadas tridimensionais dos objectos a partir da análise da evolução das suas coordenadas bidimensionais ao longo de uma sequência de imagens. O processo será levado a cabo em várias fases. Em primeiro lugar, será necessário analisar cada imagem para se detectarem certas características nela presentes que permitam identificar a posição dos objectos. De seguida, será necessário comparar duas imagens consecutivas para se determinar a nova posição de cada objecto, ou seja, estabelecer uma correspondência entre as coordenadas de cada objecto na primeira imagem e as suas coordenadas na segunda imagem. Por fim, tendo em conta as correspondências estabelecidas, será efectuada uma inferência acerca da estrutura tridimensional dos objectos presentes na cena.

O ramo da ciência que se dedica a este estudo tem a designação de Visão por Computador. Esta problemática já se encontra abordada há algum tempo. No entanto, só recentemente se tem procurado utilizar as redes neuronais para a resolução deste tipo de problemas. A sua utilização tem por finalidade tentar tirar partido da sua potência computacional na resolução de certos tipos de problemas. De facto, o seu processamento

em paralelo, a sua capacidade de aprendizagem e a sua adaptabilidade a novas situações são características muito atraentes que lhes conferem um estatuto privilegiado no domínio da Inteligência Artificial.

No capítulo seguinte, definem-se conceitos importantes ligados às TI e analisam-se os seus vários componentes, bem como também os seus princípios fundamentais e as suas funções. De seguida, explicamos de uma forma suncinta no que é que consiste a RV, quais os objectivos da sua utilização e quais os benefícios que daí poderão ser retirados. Também é efectuada uma pequena síntese da evolução histórica daqueles que podem ser considerados comos os antecessores da RV e são apresentados alguns campos de aplicação para a RV. Por fim, seão analisados alguns dos efeitos positivos que a utilização da RV poderá proporcionar às UE.

No terceiro capítulo analisamos a problemática ligada ao campo da Visão por Computador. Esta tarefa é levada a cabo através da apresentação de algumas contribuições de vários investigadores para a resolução deste tipo de problemas. A análise é estruturada tendo em conta as três fases principais na Visão por Computador atrás referidas. Antes disso, efectuamos uma apresentação sobre a evolução histórica da Visão por Computador.

O quarto capítulo é dedicado à aplicação das redes neuronais à resolução do problema da obtenção da estrutura tridimensional dos objectos a partir da análise de imagens bidimensionais. Em primeiro lugar, faz-se um enquadramento do problema. De seguida, apresentamos os modelos baseados em redes neuronais a utilizar para a resolução do problema.

Por fim, no quinto capítulo, apresentamos algumas conclusões e considerações sobre a utilização das redes neuronais na resolução do problema da obtenção da estrutura tridimensional dos objectos e acerca dos assuntos abordados neste trabalho.

2. A Realidade Virtual e as Modernas Tecnologias de Informação e Sistemas de Informação

Este capítulo tem por finalidade, em primeiro lugar, precisar o que se entende por Tecnologias de Informação, quais os seus principais componentes e quais os seus princípios e funções, e veremos também o que se entende por Sistemas de Informação, de acordo com **Senn (1996)**. A nossa preocupação será estabelecer sempre uma relação entre estes conceitos e a RV. De seguida, mostramos no que consiste a RV e, de uma forma geral, mostramos o que ela apresenta de inovador em relação às Tecnologias de Informação tradicionais e quais são as suas vantagens. Também efectuamos um breve levantamento daquilo a que se pode chamar os antecedentes históricos da RV. Para além disso, também descrevemos alguns campos de aplicação desta nova Tecnologia de Informação. Por fim, analisamos alguns dos efeitos previsíveis que a RV poderá exercer sobre as UE quando estas começarem a utilizá-la como tecnologia de informação.

2.1. Breve Introdução às Tecnologias de Informação e Sistemas de Informação

Hoje em dia vivemos numa sociedade em que a informação desempenha um papel fundamental e constitui um recurso precioso. Tal facto deu origem a uma designação apropriada para esta época: a Era da Informação. A Era da Informação teve o seu início em 1957, devido ao facto de ter sido nessa data que o número de trabalhadores envolvidos na criação, distribuição e aplicação de informação superou o número de trabalhadores pertencentes à agricultura e à indústria nos Estados Unidos.

A Era da Informação pode ser distinguida das Eras anteriores (Agrícola e Industrial) através das seguintes características:

- A Era da Informação teve a sua origem na ascensão de uma sociedade baseada na informação.
- Os negócios na Era da Informação dependem da tecnologia da informação para serem levados a cabo.

- Na Era da Informação, os processos de trabalho são transformados para se aumentar a sua produtividade.
- O sucesso na Era da Informação é determinado, em larga medida, pela eficácia com que se utiliza a tecnologia da informação.
- Na Era da Informação, a tecnologia da informação é incorporada em muitos produtos e serviços.

Portanto, as TI desempenham um papel fulcral na sociedade em que vivemos. Convém então especificar-mos o que se entende por TI.

O termo **Tecnologias de Informação** refere-se a um conjunto de items e capacidades utilizados na criação, armazenamento e dispersão de informação. As TI encontram-se divididas em três grandes componentes: computadores, redes de comunicações e *know-how*. Estes componentes serão analizados em maior pormenor nos sub-capítulos seguintes.

Tendo em conta o que foi referido no capítulo anterior, temos a opinião de que a RV se enquadra perfeitamente nesta definição. De facto, o ponto forte da RV consiste na criação e/ou modificação de informação por forma a torná-la mais facilmente apreendida pelo utilizador. Para além disso, também apresenta a possibilidade de armazenar toda essa informação e disponibilizá-la para uma série de utilizadores.

Um outro aspecto importante nas TI é o conceito de sistema. Em termos gerais, um **sistema** é um conjunto de componentes que interagem por forma a alcançarem uma finalidade. Em particular, um **Sistema de Informação** é um sistema que permite a troca de dados e informação entre pessoas ou departamentos. Um Sistema de Informação estabelece a ligação entre os outros sistemas de uma UE, permitindo-lhes trabalhar eficientemente na direcção do mesmo objectivo.

2.1.1. Computadores

Em termos muito simplificados, um **computador** é um sistema electrónico capaz de receber instruções para aceitar, processar, armazenar e apresentar dados e informação. O computador, aliado à importância crítica da informação, tem vindo a aquirir um papel fulcral na nossa sociedade como meio privilegiado para a obtenção dessa mesma

informação. Por outro lado, tem permitido a automatização de certas tarefas, conduzindo a elevados ganhos a vários níveis (rapidez, produtividade, eficiência, redução de custos, etc). Hoje em dia, encontra-se difundido em praticamente todos os aspectos da nossa sociedade, podendo nem sempre corresponder à imagem que nós temos deles. Por exemplo, encontram-se presentes nos carros, nos electrodomésticos, nas câmaras de filmar, nos semáforos, nos telefones, etc.)

Os computadores podem ser classificados em quatro tipos no que diz respeito à sua capacidade de processamento: microcomputadores, minicomputadores, *mainframes* e supercomputadores.

Os **microcomputadores** (ou computadores pessoais) são o tipo de computador mais utilizado. São relativamente compactos e de pequena dimensão. São normalmente utilizados para processamento de texto, contabilidade e finanças pessoais ou de negócios, correio electrónico, etc.

Os **minicomputadores e *mainframes***, mais potentes que os microcomputadores, estão normalmente associados a negócios de grandes dimensões, sendo utilizados para conectarem pessoas e grandes quantidades de informação. Os minicomputadores são utilizados normalmente para efetuarem tarefas específicas, ao contrário dos *mainframes*, que são utilizados para a execução simultânea de várias tarefas uma vez que são mais potentes que aqueles.

Os **supercomputadores** são o tipo de computador mais potente, tendo sido concebidos para a resolução de problemas que exijam cálculos complexos e demorados. Os cientistas utilizam-nos, por exemplo, para previsões meteorológicas, para cartografar as superfícies de planetas, na elaboração de modelos de sistemas químicos e biológicos, etc. Nos negócios, são utilizados para criar e testar novos processos, produtos e máquinas.

A RV, como tecnologia de ponta que necessita de efectuar representações tridimensionais da informação, é indissociável dos computadores. Ela só atingiu o desenvolvimento que apresenta hoje devido aos avanços na computação. Como tal, para que consiga atingir níveis mais elevados de realismo e de *performance*, está dependente dos próximos avanços na capacidade de processamento dos computadores. Para além disso, para se tornar uma tecnologia de massas, será também necessário esperar que os preços dos computadores suficientemente potentes baixem para um nível acessível à maioria dos potenciais utilizadores. Outro aspecto importante consiste na necessidade de

aperfeiçoar os periféricos existentes ou criar novos periféricos por forma a proporcionar experiências mais realistas.

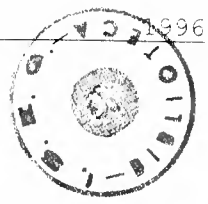
2.1.2. Redes de Comunicações

A capacidade de comunicar constitui um aspecto fundamental das TI. Tal processo é efectuado através do envio de dados e informação recorrendo a redes de comunicações. Assim, uma **rede de comunicações** consiste numa ligação entre vários locais através de um meio que permite às pessoas enviarem e receberem dados e informação. A rede mais utilizada é formada pelas infra-estruturas telefónicas, as quais, para além de permitirem estabelecer ligações entre pessoas, também conectam computadores.

A RV também recorre às redes de comunicações. Muitas das suas aplicações assim o exigem. Por exemplo, no domínio da Telepresença, existe uma necessidade de troca à distância de dados e informação entre o sistema remoto e o utilizador. Um bom exemplo disso é o caso de reuniões virtuais entre indíviduos que se encontram fisicamente separados por uma grande distância. Para que vários utilizadores possam partilhar uma experiência virtual é necessário que exista uma rede a conectar os vários nodos do sistema. Ou então, temos o caso de aplicações que envolvem a recepção de informação proveniente de vários pontos, a qual será acumulada num local em particular para um posterior processamento e análise. Portanto, a RV e as redes de comunicações são indissociáveis.

2.1.3. Know-How

Sem *know-how* é muito difícil retirarmos quaisquer proveitos das oportunidades que as TI nos proporcionam. Para que as TI sejam úteis na resolução de problemas é necessário saber reconhecer em que situações podem ser utilizadas e como devem ser utilizadas. Então, em termos gerais, ***know-how*** será a capacidade de se realizar determinada tarefa. O *know-how* inclui:



- Familiaridade com as ferramentas das TI.
- Aptidões necessárias para a utilização destas ferramentas.
- Compreender quando deve ser utilizada uma TI para resolver determinado problema ou capitalizar uma oportunidade.

É importante ser-se capaz de saber o que as TI são capazes de fazer por nós, hoje, ao mesmo tempo que se analisa o que poderão fazer no futuro. Os seus benefícios provêm de sabermos o que podemos fazer e obter através da sua utilização.

No caso da RV isso também é verdade. Há que saber analisar e identificar as oportunidades para a sua aplicação para tirar partido das suas potencialidades. Como tecnologia que se encontra no seu estádio inicial de desenvolvimento, aínda não lhe foi feita uma delimitação aplicacional. Existem muitas possibilidades em aberto que poderão ser exploradas.

Por outro lado, normalmente, um sistema de RV não exige grande *know-how* para a sua utilização. Isso deriva do facto de serem concebidos com a preocupação de permitirem uma fácil "navegação" através da informação por parte do utilizador. Para além disso, também incorporam a preocupação de permitir uma percepção mais intuitiva da informação, o que torna a sua utilização mais simples.

2.1.4. Princípios e Funções das Tecnologias de Informação

Um princípio consiste numa regra fundamental, uma linha de orientação que, quando aplicado, produzirá um resultado desejado. Os princípios não se destinam a situações em particular, a sua finalidade é servirem de orientação numa variedade de situações.

O princípio mais importante das TI também descreve a sua finalidade: "resolver problemas, desbloquear a criatividade e tornar as pessoas mais eficientes do que seriam sem a utilização das TI."

Um outro princípio importante diz que "quanto mais recorremos às tecnologias avançadas, como a tecnologia da informação, mais importante se torna considerar o seu lado humano."

O seguinte princípio encontra-se relacionado com o anterior: "devemos sempre adaptar a tecnologia de informação às pessoas e não pedir-lhes que se adaptem à tecnologia de informação."

Estes princípios sugerem que quanto mais dependermos das TI, mais importante será assegurarmo-nos de que não nos esquecemos do elemento humano.

As TI devem desempenhar seis funções relativamente à informação: captação, processamento, geração, armazenamento, recuperação e transmissão. O impacto e resultados obtidos com as TI dependem da forma como estas funções são aplicadas.

Captação

A captação de informação consiste no processo de compilar registos detalhados de actividades para serem utilizados posteriormente.

Processamento

O processamento é a actividade que é usualmente associada aos computadores. A função de processamento envolve a conversão, análise, computação e síntese de todas as formas de dados ou informação. Em particular, consiste na actividade computacional que envolve o processamento de qualquer tipo de informação e a sua transformação noutro tipo de informação.

Geração

Geralmente, as TI são utilizadas para gerar informação através do processamento. A geração de informação significa organizar dados e informação numa forma útil, quer seja como números, texto, imagens ou sons.

Armazenamento

Consiste no processo pelo qual o computador guarda dados e informação, num certo suporte, para uma utilização posterior.

Recuperação

A recuperação é o processo pelo qual o computador localiza e copia dados ou informação armazenados para processamentos adicionais ou para os transmitir a outros utilizadores.

Transmissão

A transmissão consiste no envio de dados e informação de um local para outro. Para efectuarem a transmissão, os computadores podem utilizar redes, linhas telefónicas, satélites e fibras ópticas.

Em princípio, qualquer computador é capaz de desempenhar qualquer uma destas operações, mesmo que seja necessário recorrer a algum tipo de periférico. Uma vez que a RV recorre aos computadores para ser implementada, qualquer sistema de RV será capaz de realizar estas seis funções associadas às TI.

2.2. O que se Entende por Realidade Virtual

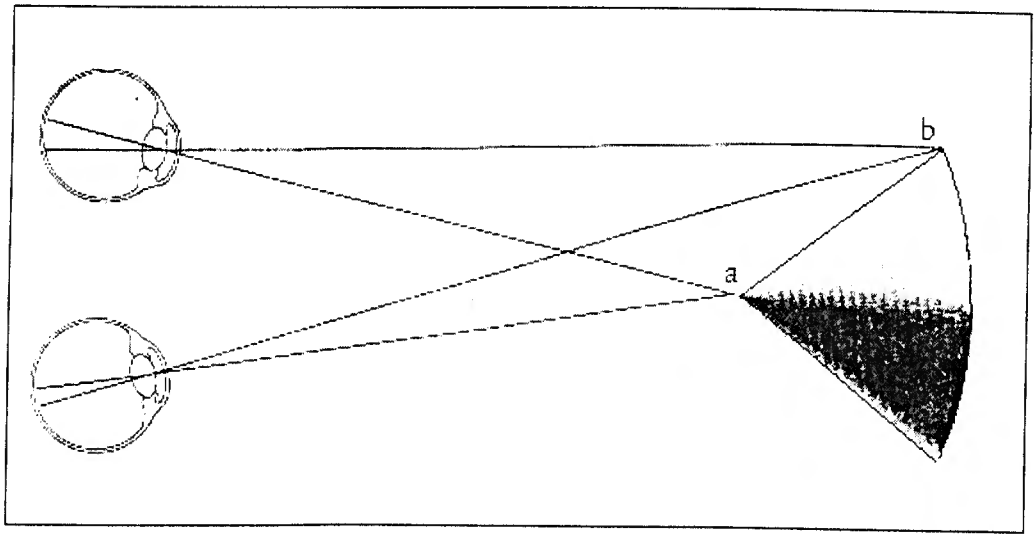
Que papel vem a RV desempenhar no contexto das Tecnologias de Informação? Podemos afirmar que a RV vem proporcionar uma forma revolucionária de percepcionar a informação. Esta nova tecnologia potencializa a percepção e compreensão da informação, aumentando a eficiência com que esta é assimilada pelo cérebro. Tal facto é alcançado através da eliminação, não só de informação de carácter não essencial, como também de uma série de factores que poderão diminuir a capacidade de concentração do indivíduo. Por outro lado, a informação verdadeiramente relevante será apresentada de uma forma destacada para que o indivíduo se aperceba da sua importância de uma forma praticamente intuitiva.

Como é que a RV possibilita tais ganhos ao nível da percepção? A sua abordagem revolucionária consiste em transformar todo o conjunto de informação numa representação tridimensional, na qual, através de um conjunto de *interfaces* específicas, o utilizador terá a sensação de se encontrar imerso. Assim, para o utilizador, a informação aparentará ter uma existência física e existir num espaço próprio. Ao utilizador fica aberta a possibilidade de "navegar" pela informação, perspectivando-a de acordo com as suas preferências. Ainda mais importante do que isso, fica-lhe igualmente aberta a possibilidade de interagir com essa mesma informação, através da manipulação da sua representação tridimensional.

Atrás, acabamos de referir duas características importantes que definem um dos tipos de RV: a **imersão** e a **interacção** com o meio virtual. A imersão visual é

conseguida mediante a utilização de um capacete que incorpora dois pequenos visores, um que projecta uma imagem destinada ao olho esquerdo e outra destinada ao olho direito. Se entre as duas imagens existir uma diferença no ângulo de visão (para simular o facto de estarem a ser captadas cada uma pelo seu olho respectivo), então o indivíduo terá a sensação de que o que está a ver tem uma existência tridimensional. Tal facto resulta da fusão das duas imagens levada a cabo pelo cérebro. Para determinar a profundidade de determinado objecto, ambos os olhos convergem e fixam esse objecto, sendo a sua imagem projectada no centro de ambas as retinas. As disparidades binoculares (diferença na posição de determinado ponto de uma imagem para a outra) resultantes do facto de se observar a mesma cena a partir de dois ângulos diferentes é que permitem obter a sensação de profundidade. Um só olho não seria suficiente para nos apercebermos de forma tão intuitiva da tridimensionalidade do mundo à nossa volta. Esse processo encontra-se ilustrado na Figura 2.1.

Figura 2.1 - O processo da visão estereoscópica



Nota: Extraído de **Poggio (1984)**

Para além da visão, também é possível criar uma sensação de tridimensionalidade para a audição. Para tal, bastará a utilização de auscultadores, os quais fornecerão sons que aparentarão vir de vários pontos no espaço tridimensional em redor do indivíduo.

Quanto ao tacto, também existem alguns dispositivos capazes de fornecer *feedback* sensitivo, sendo umas luvas próprias para essa finalidade o meio mais

utilizado. Quando o utilizador toca um objecto virtual, a luva produzirá um efeito sensitivo que dará a impressão ao utilizador de estar a tocar num objecto com existência física. Para além desse *feedback* sensitivo, normalmente, as luvas também são utilizadas como dispositivo de interacção com os objectos virtuais, permitindo ao utilizador agir sobre o ambiente que o rodeia.

Portanto, a utilização destas tecnologias permite dar ao utilizador uma sensação de total imersão num ambiente virtual, com o qual poderá interagir, se assim o desejar.

Para além deste tipo de RV, temos aínda outros dois tipos importantes que convém referir. Em primeiro lugar, temos a **Realidade Ampliada**. Esta tecnologia baseia-se na utilização de um par de óculos transparentes, que não impedem o utilizador de ver o que o rodeia, mas que permitem sobrepôr informação adicional ao que ele está a ver na realidade. Trata-se de algo semelhante ao que já se passa nos modernos aviões militares, nos quais o piloto tem um visor transparente à sua frente que lhe fornece informação relevante para a pilotagem. Esta informação aparece sobreposta à visão que ele tem do que se passa à sua frente, permitindo-lhe concentrar-se melhor na pilotagem.

Em segundo lugar, temos a **Telepresença**. A Telepresença consiste na possibilidade de o utilizador poder interagir num meio no qual ele não está fisicamente presente. O utilizador recebe à distância, através de dispositivos adequados, uma representação, tridimensional ou não, do ambiente com o qual pretende interagir. Mediante essa informação recebida, irá desencadear acções para obter os efeitos desejados. Por sua vez, essa informação será enviada para o equipamento que se encontra nesse local, o qual irá desencadear as acções pretendidas. Conforme os dispositivos utilizados, o utilizador poderá estar virtualmente imerso nesse ambiente ou não.

Tendo em conta as considerações atrás efectuadas, estamos em condições de dar uma possível definição para a RV. Assim, a nosso ver, a **Realidade Virtual** consiste num conjunto de tecnologias de comunicação e de tratamento da informação que, através de estimulações sensoriais de vária ordem, tem por finalidade proporcionar ao seu utilizador uma forma de percepcionar e interagir com essa mesma informação de uma forma mais intuítiva. Basicamente, consiste na simulação de um ambiente tridimensional representativo da informação a analisar.

Como nos apercebemos facilmente, a RV é uma Tecnologia de Informação com enormes potencialidades. No entanto, encontra-se aínda a dar os primeiros passos. Actualmente, existem dificuldades de vária ordem, principalmente no campo tecnológico. Os equipamentos utilizados na RV ainda não apresentam um nível de desenvolvimento capaz de proporcionar um grau óptimo de realismo (ao nível do processamento de imagem e de *feedback* sensitivo).

Por outro lado, dado o facto de os equipamentos diponíveis aínda se encontrarem numa fase de permanente desenvolvimento, aínda não é possível a sua produção em massa, possibilitando a diminuição do seu preço para níveis mais acessíveis aos utilizadores em geral.

Contudo, apesar destas considerações, é de prever que a RV cause um grande impacto em vários domínios da sociedade. Hoje em dia, já é estabelecida uma analogia entre o impacto previsto para a RV na nossa sociedade e o impacto provocado pela introdução da televisão na década de 60. À semelhança do que aconteceu com a televisão, a RV poderá traduzir-se numa nova forma de percepcionar e compreender o mundo que nos rodeia, ampliando a extensão e a eficiência dos nossos sentidos. À RV caberá o papel de funcionar como *interface* entre o mundo humano e o mundo dos computadores, permitindo ao seu utilizador compreender de forma mais intuítiva a informação disponível por meio das "auto-estradas de informação".

Estamos, portanto, perante uma tecnologia singular e com princípios inovadores, que promete revolucionar as tecnologias de informação, abrindo-se inúmeros campos de aplicação. No sub-capítulo 2.4 daremos alguns exemplos de campos de aplicação para a RV.

2.3. Síntese Histórica da Realidade Virtual

A seguinte análise histórica baseia-se na pesquisa desenvolvida por **Pimentel e Teixeira (1993)**. Esta obra também será utilizada como referência para o sub-capítulo 2.4.

Os princípios subjacentes à RV podem-se identificar com a tentativa de encontrar formas ou técnicas de capturar a essência de uma experiência, permitindo a

disponibilização da informação para um conjunto de observadores. As raízes deste tipo de preocupação remontam ao início do século XV. Por volta de 1400, Giotto, um pintor Florentino, criou um método intuítivo para projectar figuras tridimensionais em quadros de duas dimensões. Esta descoberta foi inovadora no sentido de ser a primeira a tentar transmitir a sensação de profundidade através do uso da perspectiva, tendo sido largamente utilizada por artistas de gerações seguintes.

Ao longo dos anos, as tentativas de armazenar e recriar experiências avança continuamente no sentido de maximizar o realismo, sendo de destacar, por exemplo, a pintura efectuada em 1788 por Robert Baker, em que este artista escocês representava a cidade de Edimburgo numa tela invulgar que cobria os 360° graus de visão do observador. Esta pode ser considerada a primeira representação verdadeiramente envolvente, dado que todo o campo de visão do espectador é absorvido pela representação. Mais uma vez, este tipo de representação teve larga difusão no meio artístico.

Por meados de 1830, a recém criada tecnologia da fotografia torna-se bastante popular, permitindo um nível de realismo nunca conseguido anteriormente pela técnicas tradicionais. Neste contexto, surge em 1833 uma técnica revolucionária capaz de elevar o realismo da representação a níveis bem mais elevados, a projecção estereoscópica. Esta técnica de representação, criada por Wheatstone e melhorada em 1844 por David Brewster, utilizava dois fotogramas da mesma imagem, com a particularidade de um dos fotogramas ter sido captado de um ponto de observação ligeiramente distanciado do do outro fotograma. Com tal procedimento, e utilizando uma técnica especial, cada um dos fotogramas é colocado junto do olho correspondente, reproduzindo o processo através do qual o nosso cérebro se apercebe da profundidade dos objectos presentes numa cena. Estavam assim assimilados os rudimentos da visão estereoscópica e a sua relação com o nível de percepção.

O passo seguinte no sentido da simulação da realidade foi conseguido através da invenção da imagem em movimento: o cinema. A projecção, em 1895, do primeiro filme dos irmão Lumière traduziu-se em algo completamente inovador. A dinâmica da imagem em movimento tinha sido reproduzida. Daqui em diante, o aparecimento do cinema sonoro, a introdução da película fotográfica colorida, surgem como evoluções no sentido de proporcionar uma experiência mais realista.

Com o aparecimento das primeiras emissões de televisão, por volta de 1941, foi acrescentada uma nova dimensão no domínio da representação, a telepresença, a sensação de se estar presente no local da câmara. O facto de a televisão permitir transmitir imagem e som em directo a espectadores distantes possibilita a experimentação em tempo real, facto que não era possível através do cinema.

No campo da reprodução visual e sonora, foram posteriormente desenvolvidas algumas técnicas para aumentar o nível de realismo, sendo dignos de destaque algumas que alcançaram bastante popularidade. Por exemplo, na década de 50, uma tentativa de retomar a sensação de uma experiência envolvente através de uma variante do cinema convencional: o Cinerama. O Cinerama consistia em proporcionar ao espectador uma reprodução quase óptima do ponto de vista visual e sonoro. Para isso, utilizava a projecção de um filme especial (três câmaras sintonizadas de 35 mm), que abrangia a totalidade do ângulo de visão do espectador. Ao nível do som, eram utilizadas seis pistas de som estereofónico que garantiam um realismo sonoro nunca alcançado anteriormente. Este sistema, apesar de ter provado o interesse dos indivíduos em experimentar experiências de elevado grau de realismo, não teve continuidade devido ao elevado custo associado à sua produção e manutenção tecnológica.

Por fim, e como expoente máximo dos antecessores da RV, surge em 1960, por intermédio do americano Morton Heilig, o qual ficou fascinado com o Cinerama, a primeira máquina destinada a reproduzir uma experiência integrando uma variada gama de sentidos (visão, audição, olfacto e tacto). É, por estas características, uma abordagem revolucionária, apresentando pela primeira vez a integração simultânea da visão estereoscópica, som estereofónico, simulação de odores e reprodução de movimento. Ao dispositivo capaz de proporcionar tal envolvente experiência foi dado o nome de Sensorama. Formalmente, era um dispositivo com uma aparência semelhante às actuais máquinas de diversão, apresentando um dispositivo binocular (para a visualização de imagens estereoscópicas), um sistema de altifalantes, um mecanismo de libertação de odores colocado perto do nariz e aínda um mecanismo vibratório integrado com pequenas aberturas de ventilação para transmitir a sensação de movimento ao observador. Esta máquina, embora considerada a melhor tentativa tecnológica de recriação da realidade, não obteve sucesso comercial porque os meios tecnológicos utilizados eram demasiado dispendiosos para que a sua difusão fosse lucrativa em termos comerciais.

Nas duas últimas décadas, o progresso tecnológico tem sido verdadeiramente notável, principalmente no campo da Informática. O computador tem vindo a ganhar um papel essencial e destacado na sociedade actual. Em resultado da intensa concorrência e da sua massificação comercial, o computador adquiriu um papel privilegiado como ferramenta de trabalho e de lazer junto do cidadão comum e na sociedade em geral. É difícil encontrar um aspecto da vida em sociedade em que o computador não tenha sido ou não possa vir a ser utilizado como instrumento de dinamização da informação.

A RV, encontrando-se directamente dependente da tecnologia, surge-nos completamente interligada à tecnologia computacional. Na verdade, a divulgação que actualmente se nota no domínio da RV deriva do facto de o potencial tecnológico presente nos computadores pessoais comuns ter atingido uma *performance* muito elevada, sendo capazes de responder aos requisitos impostos por uma simulação de qualidade.

Como o desenvolvimento da RV se encontra intimamente ligado às tecnologias da computação, pareceu-nos lógico fazer uma interligação com a tecnologia das redes neuronais, as quais, como iremos mostrar mais à frente, constituem um meio privilegiado para certos tratamentos computacionais da informação. Daí, a razão deste trabalho, o qual sugere uma interligação possível para estas duas jovens e promissoras tecnologias.

2.4. Campos de Aplicação da Realidade Virtual

Depois de termos abordados os princípios básicos da RV e os seus antecedentes históricos, vamos agora apresentar alguns campos em que ela já está ou poderá vir a ser aplicada. Os campos de aplicação, dadas as suas grandes potencialidades no tratamento e comunicação da informação, são vastos e complexos, como se poderá concluir pela breve descrição de alguns deles.

2.4.1. Aplicações militares

A simulação militar é um dos campos de aplicação em que a RV é utilizada há mais tempo. Há muito que os responsáveis militares se aperceberam que a simulação constitui um activo precioso. Permite sujeitar os soldados a situações que representam uma simulação da realidade, proporcionando um treino sem a necessidade de pôr em risco vidas humanas e material de guerra dispendioso. A simulação fornece-lhe conhecimentos valiosos que serão postos em prática quando enfrentarem uma situação real. A sua rapidez de reacção e o seu discernimento serão muito mais eficazes.

As forças armadas dos EUA possuem um gigantesco simulador militar, a SIMNET. Trata-se de uma rede de telecomunicações que permite fazer a ligação entre vários postos de simulação, possibilitando uma interacção num ambiente comum a todos eles. As vantagens da utilização de simuladores interligados com redes são enormes. Permite que vários soldados possam interagir, ou seja, poderão agir em equipa e aprender a combater lado a lado, a funcionar eficientemente em conjunto. Isso é extremamente importante, pois deve existir uma sincronia entre todas as partes participantes em manobras militares.

É esse o princípio que está subjacente a um novo projecto, a Inter-rede de Simulação Distribuída (DSI). Trata-se de um projecto ambicioso, através do qual se pretende interligar dez mil simuladores. Será uma rede global de simulação em tempo real, através de fibra óptica e assistida por satélite. Obter-se-á uma total coordenação entre os vários ramos das forças armadas utilizando um protocolo comum a todos.

Na sequência da Guerra do Golfo, o Pentágono ficou com uma representação digital da zona na qual se desenrolou a crise: Arábia-Saudita, Koweit e Iraque, dando origem à base de dados SAKI. A zona encontra-se mapeada metro a metro, em três dimensões e com opções para várias condições atmosféricas. A 26 de Fevereiro de 1991, no deserto a Sul do Iraque, verificou-se um combate entre um regimento de cavalaria blindada dos EUA e uma divisão da guarda republicana do Iraque. Apesar de as tripulações americanas não terem experiência anterior de combate, enquanto que as forças iraquianas eram compostas pela elite de veteranos da guerra de oito anos entre o Iraque e o Irão, os americanos, sem qualquer apoio aéreo, aniquilaram os iraquianos em menos de 22 minutos. Esta batalha foi o combate que se registou com maior precisão e detalhe em toda a história. Os participantes americanos foram exaustivamente

entrevistados e o terreno de batalha foi medido metro a metro. O resultado foi a obtenção de uma réplica digital, interactiva e oferecendo a opção de ligação em rede. Todos os factos estão lá representados. Os seus utilizadores podem agora observar os acontecimentos tridimensionalmente, de qualquer ângulo, durante qualquer momento do combate.

Como exemplo da utilização da Telepresença na Guerra do Golfo, refira-se a utilização de *drones*, na forma de aviões não tripulados. Estas aeronaves são controladas à distância, sendo utilizadas para se obter imagens das posições do inimigo. O operador do sistema pode desencadear um bombardeamento sobre a posição observada na imagem.

A tecnologia de simuladores encontra-se tão desenvolvida ao ponto de as fotografias de satélite poderem ser automaticamente transformadas em representações virtuais e armazenadas em bases de dados para serem posteriomente utilizadas em treinos. Assim, antes de participarem em qualquer conflito, em qualquer parte do globo, os soldados já conhecerão o terreno palmo a palmo e já terão levado a cabo centenas de missões de treino. Poderão chegar ao ponto de conhecerem o terreno melhor que o próprio inimigo.

A RV também poderá ser utilizada para simular e testar armamento antes mesmo de este ter sido construído, permitindo poupar fundos em testes reais. Este armamento virtual até poderá tornar-se obsoleto antes mesmo que se chegue a produzir uma única unidade.

2.4.2. Aeronáutica

A Força Aérea dos EUA está a estudar a possibilidade de utilizar a RV para facilitar o trabalho dos controladores de tráfego aéreo. A intenção é explorar uma forma de colocar os controladores no espaço aéreo, juntamente com os aviões. O controlador, em vez de estar sentado em frente a um monitor, encontrar-se-á posicionado por cima do aeroporto, tendo uma visão global da situação. Será possível monitorar e comunicar com os aviões baseando-se na sua posição num espaço a três dimensões. Ver-se-á um modelo do aeroporto e os aviões que são da responsabilidade desse controlador em particular encontrar-se-ão claramente identificados. Para além disso, poder-se-ão visualizar outras

actividades que ocorram na área do aeroporto. Será possível avaliar rapidamente a distância e a altitude de um objecto apenas olhando para ele e qualquer informação adicional poderá ser obtida apenas com uma simples palavra de comando.

Suponhamos que acontece uma situação de um eminente desastre em que um avião se prepara para aterrar numa pista na qual já se encontra um outro procedendo a uma descolagem. Como convém neste tipo de situações, o computador fará desaparecer tudo o que não seja essencial e que potencialmente possa distrair a atenção. Rotas de escapatória serão iluminadas e quaisquer veículos ou aviões na vizinhança permanecerão visíveis

Este é um bom exemplo do que diferencia a RV dos outros meios de comunicação de informação: a possibilidade de decidir o que deve ser representado, onde deve ser representado e o que não deve ser representado. Ou seja, existe uma grande interactividade que nos permite personalizar a informação face às nossas necessidades.

Uma outra possibilidade de aplicação da RV à aeronáutica situa-se no campo do *design*. Hoje em dia, os *designers* podem modelar e simular as suas idéias, em duas dimensões, nos écrans de computadores utilizando programas de CAD. No entanto, com a aplicação da RV, poderão "colocar as suas mãos dentro do écran" e entrar num espaço virtual no qual procederão à modelização. Por exemplo, a Northrop está a utilizar um sistema de RV para actualizar o *design* do caça F-18 da Força Aérea americana.

A Boeing já se encontra a utilizar esta nova tecnologia, desenvolvendo activamente uma grande pesquisa neste campo. Por um lado, está a estudar a utilização da Realidade Ampliada, principalmente para aplicação a operações de manufacturação. O objectivo é aumentar a produtividade dos trabalhadores, proporcionando-lhes a informação de que necessitam, na altura necessária. Esta aplicação será importante para o pessoal da manutenção. Com a utilização de óculos de RV transparentes, poderão obter informação preciosa sem terem de retirar as mãos do seu trabalho. A informação necessária será colocada na retina do utilizador sem que ele tenha que parar a realização da sua tarefa. Isto será especialmente útil nos casos em que os trabalhadores não possuam treino suficiente ou quando o modelo do avião não seja familiar.

O equipamento de RV até poderá fornecer uma espécie de visão Raios-X. Em muitas reparações de motores a jacto, os mecânicos não podem ver as suas mãos pois elas estão dentro do motor. Com a utilização de uns óculos de RV transparentes e através

da monitorização da posição das mãos em relação às peças, o mecânico poderá "ver através" da maquinaria até ao ponto onde as suas mãos estão a trabalhar.

Um outro grupo de trabalho está a estudar a imersão total em ambientes de RV para integração nos processos de *design*, testes e obtenção de modelos experimentais. Existe a intenção de utilizar esta tecnologia para procederem ao *design* de uma nova geração de aviões comerciais, os 777.

2.4.3. Medicina

Actualmente, a medicina já recorre a uma série de tecnologias avançadas (endóscopia, cirúrgia laser, ultra-sons, ressonância magnética, etc). A finalidade de algumas destas técnicas é a obtenção de imagens do interior do corpo humano, revelando tumores, mal-formações, hemorragias, lesões e uma variedade de patologias sem a necessidade de se proceder a uma intervenção cirúrgica.

Com a sua capacidade para gerar representações gráficas a três dimensões, a RV é uma tecnologia que poderá trazer importantes avanços a este campo. A RV poderá proporcionar novas ferramentas para a investigação, formação e tratamento.

No campo da investigação, a RV está a auxiliar o desenvolvimento de novos medicamentos. Os bioquímicos estão a utilizá-la para melhor compreenderem a estrutura e propriedades das moléculas orgânicas através da sua "manipulação" tal como se elas se encontrassem fisicamente em frente a eles.

Também se está a testar a sua aplicação como auxiliar nos diagnósticos. Na Universidade da Carolina do Norte, está a ser desenvolvido um conjunto de óculos de RV que serão ligados a um *scanner* de ultra-sons. A imagem gerada pelos ultra-sons será sobreposta em óculos transparentes. Por exemplo, um obstetricista poderá usar este equipamento para examinar o feto de uma mulher grávida, tal como se fosse dotado de visão de Raios-X.

No campo da formação cirúrgica, algumas empresas já estão a desenvolver simulações de treino baseadas em RV. Vídeos retirados de operações cirúrgicas poderão ser integrados, em conjunto com gráficos gerados por computador, em simulações deste tipo. A sua utilização será um precioso auxiliar no processo de treino, permitindo evitar erros provocados pela falta de experiência.

No domínio das operações cirúrgicas, alguns investigadores estudam a possibilidade de aplicação da Telepresença, originando uma forma de telecirúrgia. Num futuro próximo, os cirurgiões poderão ter a possibilidade de trabalhar a partir de uma *workstation* cirúrgica. O médico e os seus assistentes prepararão o paciente, colocando os instrumentos endoscópicos e ligando-os a um sistema robótico que poderá controlar o seu movimento na direcção requerida. A partir da sua *workstation*, o médico poderá ver num monitor uma imagem tridimensional proveniente da endóscopia. Num outro monitor (ou numa janela), poderá obter uma imagem do corpo de um ângulo diferente, gerada por computador, para efeitos de posicionamento dos instrumentos. Depois, utilizará uma *interface* que lhe proporcionará *feedback* táctil para as suas mãos. Estes instrumentos fornecer-lhe-ão posicionamento e pressão à escala por forma a que o movimento das suas mãos seja transferido para os instrumentos cirúrgicos ao nível microscópico, se necessário. Informação médica importante poderá ser sobreposta no monitor para um acesso fácil ou então ser convertida em voz sintetizada, conforme a preferência. Para além disso, o cirúrgião poderia ensaiar os procedimentos momentos antes de os fazer no corpo humano. Através desta simulação, poderia achar a melhor solução antes de fazer a execução real.

Esta *workstation* cirúrgica poderia também ser utilizada, em conjunção com um microscópio de electrões, para permitir a microbiólogos e engenheiros genéticos trabalharem directamente dentro das células.

Uma outra aplicação possível para a RV na medicina seria a criação de corpos virtuais. Ao colocar os óculos, o utilizador veria à sua frente uma representação tridimensional de um corpo. Este poderia ser transparente para permitir estudar orgãos e sistemas. Um écran com informação poderia ser visualizado em qualquer local à sua escolha. Para fins de ensino, um grupo de estudantes poderia também usar óculos semelhantes. Como a simulação seria totalmente gerada por computador, os estudantes poderiam escolher qualquer ponto de vista dentro da sala de operações, inclusivé a vista que o professor estivesse a receber. Para além disso, seria possível fazer uma ampliação de vários pormenores. Também se poderia demonstrar os efeitos da administração de certos medicamentos e drogas sobre os vários orgãos e sistemas, iluminando-os à medida que aqueles se espalham pelo corpo. Com a utilização da RV também será possível efectuar viagens dentro do corpo humano, observando, do seu interior, o funcionamento de qualquer parte do corpo.

2.4.4. Arquitectura

A RV também pode ser utilizada no design e teste de edifícios. Por exemplo, no caso de edifícios destinados a escritórios ou a instalações industriais, poder-se-á utilizar um modelo tridimensional dessas instalações para se testar a funcionalidade do seu espaço e determinar se de facto se adequam à finalidade pretendida. Quaisquer correcções que seja necessário efectuar serão levadas a cabo com um custo muito menor do que aquele que seria necessário suportar para fazer essa correcções depois de finalizada a construção. Portanto, a RV permitirá uma visualização prévia dos espaços, permitindo determinar com antecedência a sua funcionalidade e determinar se permite ou não maximizar a produtividade dos trabalhadores.

Para além disso, a RV também poderá ser útil para arquitectos que concebam edifícios para habitação. Por exemplo, a Matsushita Electric Works concebeu um modelo virtual de uma cozinha que permite aos seus clientes potenciais navegarem pelas várias cozinhas-modelo concebidas pela sua divisão de construção de moradias. O resultado é que a Matsushita pode conceber no próprio momento uma cozinha à medida do cliente e obter a sua aprovação em apenas 30 minutos. Para além disso, verificaram que este sistema provocou um aumento no seu volume de negócios, uma vez que também funciona muito bem como ferramenta promocional.

2.4.5. Mercado de títulos

Na actualidade, a actividade na área da Gestão caracteriza-se pela necessidade de utilizar grandes quantidades de informação. Para que se possa fazer uma gestão eficiente de qualquer UE, é necessário recorrer a uma grande quantidade de fontes de informação e acompanhar a evolução de inúmeras variáveis importantes.

Isso é particularmente importante no caso do mercado de títulos, em que os operadores têm que acompanhar várias fontes de informação, das quais estão a receber dados em tempo real. Para além de terem que monitorar toda esta informação, também têm que efectuar decisões rápidas das quais poderão resultar avultados lucros ou prejuízos.

A aplicação da RV proporcionará uma forma de aumentar a eficiência dos operadores. Isso será conseguido através da condensação e uniformização da forma como a informação será por eles captada, permitindo-lhes a tomada de melhores decisões e de uma forma mais rápida.

Apresentamos de seguida o testemunho de Paul Marshall, presidente da Maxus Systems International, uma empresa que desenvolve sistemas para a gestão de carteiras de títulos: "A RV proporcionará o mesmo tipo de vantagem que de início obtive com a utilização da interface Windows da Microsoft, ou seja, o benefício de poder compartilhar dados com outros programas em multitarefa e poder monitorar a evolução de muitas variáveis em simultâneo. A RV será muito potente, permitir-me-á unificar informação abstracta proveniente de muitas fontes, recolher informação de cotações internacionais e de várias bases de dados e acompanhar tudo isto em três dimensões. Os *outputs* de 45 relatórios, por exemplo, poderão ser sumariados no comportamento de um polígono que muda de côr, pisca ou gira, comportando-se de uma forma pré-determinada para comunicar instantaneamente ao operador o que aconteceu a esse título."

Utilizando um equipamento de RV, Paul Marshall criou um vasto oceano de dados no qual os títulos sobem e descem nas ondas do mercado. Esta área pode ser interactivamente dividida em qualquer combinação de sub-regiões numa grelha. A grelha pode ser dividida em vários tipos de indústria, por exemplo transportes, construção, computadores, etc. Os transportes podem por sua vez ser divididos em camiões, ligeiros para passageiros, etc. Ou então, poderá ser dividida em países. O utilizador pode "voar" para cada sub-região e mover-se entre as acções e as obrigações, sendo cada uma representada por um polígono que se encontra a subir ou a descer num mar ondulante de títulos. A forma do título, a sua posição, comportamento e côr reflectirão as condições do mercado. As representações dos títulos poderão ter o logotipo da companhia a que pertencem para que sejam mais facilmente identificados.

2.4.6. Folhas de cálculo

As folhas de cálculo têm desempenhado um papel importante como ferramenta auxiliar na gestão das UE. Para além de possibilitarem o armazenamento de dados, o seu grande potencial consiste em possibilitarem simular o comportamento de mercados,

produtos, empresas, etc. Portanto, uma folha de cálculo pode ser classificada como um simulador, fornecendo uma metáfora simbólica para o comportamento de um negócio. Permite simplificar e condensar uma grande quantidade de dados complexos, podendo o utilizador fazer uso dessa informação para controlar a "saúde" do negócio e/ou criar vários cenários alternativos e ver qual o resultado que seria obtido.

A RV funciona de maneira semelhante. Permite apercebermo-nos de conceitos e idéias abstractas para as quais não existe um modelo ou representação real. A RV agirá como um tradutor, convertendo estes conceitos e idéias em experiências que os nossos sentidos poderão analisar de forma mais intuítiva.

Em organizações que apresentam uma dimensão considerável, as folhas de cálculo encontram-se instaladas em *mainframes* ou em PC´s ligados em rede. Automaticamente, recebem e armazenam informações provenientes de vários pontos da empresa. Toda esta informação armazenada pode ser utilizada para analisar a sua condição. O que falta é uma forma fácil de representar, apresentar e interpretar toda esta informação. Essa será a função da RV. Tomando essa informação abstracta, dar-lhe-á vida através de uma folha de cálculo de 2ª geração, a folha de fluxos virtual (*virtual flowsheet*). Esta folha de fluxos permitirá a utilização de várias variáveis para se criar uma representação animada e interactiva de grandes quantidades de dados. Uma empresa assemelha-se a um organismo vivo, sendo influenciada por centenas de variáveis. Se estas variáveis forem representadas por uma metáfora visual, um observador poderá ver uma representação de vários dias na "vida" da empresa em apenas alguns segundos. Poderá ver a sua evolução como entidade "viva", em constante mutação e fazer previsões acerca da sua evolução futura, tal como o fazem hoje em dia os meteorologistas para as previsões meteorológicas.

2.5. Efeitos da Utilização da Realidade Virtual nas Unidades Económicas

Comos já vimos atrás, a RV pode ser considerada como uma TI. De facto, trata-se de uma TI inovadora que apresenta conceitos revolucionários. Com a sua aplicação às UE, a forma de estar destas perante a informação deverá sofrer uma modificação assinalável.

De facto, esta nova TI, ao apresentar uma *interface* inovadora entre o utilizador e o computador, irá conduzir a importantes mudanças estruturais, culturais e de processos nas UE. Serão mudanças que permitirão obter ganhos acentuados no nível qualitativo da informação. Todos os níveis da estrutura da empresa poderão obter importantes benefícios na sua utilização através da implementação de diferentes aplicações adequadas às necessidades de cada um deles. Por exemplo, ao nível estratégico, as Políticas e Estratégias definidas poderão ser testadas num mundo virtual para se saber qual será o seu desempenho, antes de serem implementadas na realidade. Por outro lado, esse mundo virtual poderá ser utilizado para simular vários cenários alternativos, algo que pode ser fundamental para o planeamento contingencial.

Ao nível táctico, também poderão ser utilizadas simulações para se testar o desempenho dos Planos, Programas e Orçamentos. Poder-se-á também observar, virtualmente, em que medida é que cada um dos Planos e Procedimentos irão contribuir para que os objectivos traçados para a UE sejam alcançados.

Por outro lado, através de uma representação virtual da UE, com um *input* de dados e informação em tempo real proveniente de todos os seus pontos, poderemos apreender de uma forma mais fácil e clara o seu desempenho como um todo e a forma como estão a funcionar os seus sub-sistemas. Através desta representação, será fácil detectar eventuais problemas e proceder rapidamente à sua correcção. Assim, as várias áreas funcionais poderão trabalhar em síncronia, permitindo que, globalmente, o sistema seja mais eficiente e funcione como um todo.

Ao nível operacional, a RV também permitirá obter importantes ganhos. Por exemplo, no treino e/ou recrutamento de recursos humanos. A possibilidade que a RV tem de gerar qualquer ambiente permitirá a recriação, em termos virtuais, de determinado posto de trabalho. Esta simulação possibilitará ao indivíduo receber um treino realista para a tarefa a desempenhar futuramente sem ter chegado a ter qualquer contacto com o seu posto de trabalho.

Por outro lado, uma vez que a RV introduz uma forma inovadora de apresentar a informação, permitindo a sua melhor percepção, o desempenho do seu utilizador nas tarefas operacionais será muito mais elevado graças a este precioso auxílio.

Para além disso, a utilização de redes permitirá uma maior aproximação entre as UE. De facto, a RV permite a interação de vários intervenientes num ambiente virtual. Isto é ideal para a realização de encontros de negócios entre gestores sem que estes

tenham que se deslocar fisicamente para um espaço comum a todos eles. Por outro lado, permitirá também a troca de vários tipos de informação, o acesso a bases de dados virtuais, a realização de testes conjuntos, a troca de experiência acumulada, enfim, uma maior colaboração, se esse for o seu desejo.

Por tudo o que foi referido atrás, facilmente se conclui que a introdução da RV nas UE virá provocar profundas alterações nos seus Sistemas de Informação. A forma de aceder e utilizar a informação será completamente diferente, pois a *interface* a utilizar será inteiramente nova. Mais do que nunca, o utilizador estará próximo da informação, possibilitando-lhe interagir com ela de forma muito mais acentuada.

3. A Investigação no Âmbito da Visão por Computador

Neste capítulo é feita uma curta análise à evolução histórica da pesquisa nesta área e às tentativas dos investigadores para simularem o processo biológico da visão através da utilização de computadores. Também são referidos alguns campos de aplicação para os avanços alcançados na visão por computador. O corpo principal do capítulo consiste na descrição de alguns algoritmos desenvolvidos pelos investigadores para simularem a visão. Esta enumeração não pretende ser exaustiva. O seu objectivo é apenas ilustrar, através da apresentação desses algoritmos, os problemas e desafios que se têm colocado aos investigadores e apresentar algumas das soluções por eles desenvolvidas para os superarem.

3.1. Síntese Histórica

O interesse na área da visão por computador e no processamento de imagens por computador remonta ao inicio da década de 60. Nessa altura, os investigadores de Inteligência Artificial estavam confiantes de que o problema da visão seria fácil e rapidamente resolvido. Essa confiança provinha da constatação de que a visão é levada a cabo pelo sistema visual humano automaticamente e sem qualquer esforço. De facto, para obtermos informação visual sobre o mundo que nos rodeia não é necessário realizar qualquer esforço consciente. Este excesso de confiança depressa caiu por terra perante as dificuldades que se foram levantando e, nos nossos dias, a visão continua a ser um dos maiores desafios que se coloca aos investigadores da Inteligência Artificial.

Os mecanismos biológicos que nos permitem ver atestam essa complexidade. O processo inicia-se nos olhos. A retina contém mais de 125 milhões de receptores que são responsáveis pela captação de luz. Essa luz é depois transformada em impulsos nervosos, separando assim a informação útil e ignorando tudo o que seja irrelevante, funcionando como um filtro. Até que a imagem passe pelo nervo óptico, mais de 10 biliões de operações terão sido efectuadas em menos de um segundo. Posteriormente, as imagens são processadas no cortex cerebral, local onde toda a informação sensorial é

tratada. Só a visão ocupa 60% do cortex, enquanto os restantes 40% se encontram distribuídos por todos os outros quatro sentidos.

Hsu e Kusnan (1989) dão uma breve descrição histórica do caminho percorrido pela investigação na área da visão por computador.

As primeiras tentativas efectuadas para simulação da visão remontam à década de 50, com a tentativa de reconhecimento de imagens a duas dimensões, como por exemplo, os caracteres do alfabeto. A identificação era feita comparando imagens com outras imagens-padrão previamente introduzidas na memória. O reconhecimento consistia num processo estatístico de análise numérica das intensidades de luminosidade das imagens a reconhecer por contrapartida das imagens-padrão. Cada imagem introduzida era comparada a todas as imagens-padrão existentes em memória e para cada comparação era calculado um grau de semelhança. A que apresentasse um grau de semelhança mais elevado corresponderia à identificação correcta. É claro que o sistema estava limitado ao número de imagens-padrão introduzidas, o qual não poderia ser muito elevado pois o processo de comparação ficaria muito lento.

Este processo pioneiro era muito rudimentar, encontrando-se muito distante de uma simulação eficiente da visão animal. No entanto, teve o mérito de ser o precursor das primeiras tentativas efectuadas sobre a visão por computador e a interpretação de imagens a três dimensões levadas a cabo na década de 60. Foi por esta altura, quando os investigadores tentaram começar a escrever programas de computador para visão, que se aperceberam da verdadeira dificuldade do problema. A pesquisa estava principalmente orientada para a definição dos contornos (limites) de objectos, até se concluir que os contornos são uma construção da mente. As ambiguidades e subtilezas presentes nas imagens tornavam os sistemas desenvolvidos inadequados para o processo da visão. Para fazerem face a estas dificuldades, os investigadores começaram a incorporar conhecimentos nos seus sistemas de visão. A finalidade era que esses conhecimentos permitissem ao sistema eliminar os objectos que não pudessem existir em três dimensões.

Os modelos aplicados colocavam uma restrição na forma como se podiam combinar linhas, cantos e faces de objectos por forma a se obterem resultados realistas. Esta informação geométrica era então utilizada para identificação de objectos através de uma análise lógica, sem necessidade de se recorrer a comparações com imagens-padrão.

Esta aproximação de natureza cognitiva exigia uma grande capacidade de processamento, o que se tornou viável na década de 70 com a introdução do microprocessador. Estes sistemas já eram mais robustos mas apenas podiam ser aplicados a imagens simples e com figuras geométricas pouco complicadas.

David Marr, com o seu trabalho de investigação no domínio da visão por computador, marcou a década de 70. Aínda hoje os investigadores desta área vão buscar linhas directivas e inspiração para as suas pesquisas aos desenvolvimentos por ele introduzidos. Foi um dos primeiros a combinar as áreas da Inteligência Artificial e da visão com a psicologia experimental. A sua abordagem mostrou que o campo da visão constitui uma ciência por si só e a terminologia por ele introduzida tornou-se o padrão seguido neste domínio. As suas teorias foram reunidas e publicadas postumamente sob o nome *Vision*, em 1982, tornando-se numa obra incontornável.

Segundo Marr, um sistema de visão para ser bem sucedido deve, em primeiro lugar, identificar as superfícies dos objectos e as suas orientações antes de identificar os objectos propriamente ditos. Os métodos até então existentes, denominados métodos de segmentação, tentavam identificar objectos baseando-se em áreas com sombreados e intensidades (de luminosidade) idênticas. Estes métodos falhavam redondamente quando as imagens continham reflexos ou sombras. Marr sugeriu que este problema seria ultrapassado se em primeiro lugar se identificassem as superfícies.

Segundo ele, o processo da visão por computador pode ser dividido em duas fases principais:

- primeiro, numa fase inicial (visão primária ou de baixo nível), o objectivo é extrair o máximo de informação útil da imagem em bruto sem se identificar o seu conteúdo;
- depois, essa informação será combinada com conhecimentos do mundo real por forma a se determinar o conteúdo da imagem.

A imagem será introduzida no sistema de visão como um conjunto de pontos (*pixels*), cada um deles com um certo valor numa escala de cinzento (intensidade da luminosidade). Cada um desses pontos pode ser identificado pela sua posição num sistema de coordenadas cartesiano.

O primeiro passo consiste na obtenção de um primeiro esboço composto pelos contornos dos objectos e outras características da imagem com interesse. Através de um processo de filtragem, o "ruído" presente na imagem será retirado, enquanto que, simultaneamente, fará sobressair as características essenciais. A imagem resultante será muito diferente da original, apenas um esboço grosseiro. No entanto, apresenta a vantagem de ser mais facilmente trabalhada pelo computador uma vez que contém menos dados (só contém cantos, linhas, contornos, etc) e menos ambiguidades.

Assim, um dos problemas mais estudados neste campo de investigação é a detecção dos cantos e/ou contornos (*edge detection*) para a criação do esboço. Se o método for eficiente, o resultado será um conjunto de linhas que traduz os contornos e limites físicos dos objectos contidos na cena, deixando em branco as superfícies compreendidas entre eles.

Um dos métodos de detecção de contornos é um filtro baseado na curva de Gauss. Este filtro elimina o "ruído" presente na imagem através de um processo de alisamento de cada ponto em relação aos seus vizinhos. Primeiro, o filtro mancha a área a ser alisada somando valores a esse grupo de pontos. Depois, faz uma média desses pontos, assegurando-se de que pequenas mudanças na intensidade são alisadas (a intensidade mantém-se estável na superfície dos objectos, alterando-se bruscamente nos seus contornos), enquanto que as maiores variações são preservadas. Assim, o resultado será um esboço no qual deverão estar salientes os contornos.

Este primeiro processo de identificação de contornos ou de outras características de interesse é um processo de baixo nível no processamento da visão. A identificação dos objectos, da forma como ocupam o espaço e da sua estrutura tridimensional é um processo de alto nível. É este último processo que tem colocado os maiores desafios aos investigadores. Já é ponto assente que é necessário introduzir conhecimentos do mundo real nos algoritmos por forma a possibilitar o reconhecimento de objectos ao eliminar o que seja fisicamente impossível. Isto reduzirá as ambiguidades presentes na imagem e o número de interpretações possíveis. Para o conseguirem, os investigadores sabem que devem recorrer a características como texturas das superfícies, distâncias, orientação, formas e brilho. Na prática, a dificuldade está em criar a Inteligência Artificial a incorporar nos algoritmos que traduza esses conhecimentos do mundo real.

O olho humano tem células que se dedicam especialmente ao fornecimento de informação sobre a forma, orientação, textura e côr, enquanto que a visão por

computador só pode recorrer ao esboço para obter informação. Por exemplo, para a obtenção de distâncias num sistema de visão estereoscópica, será necessário comparar os esboços esquerdo e direito de uma cena. Essa comparação permitirá identificar pontos equivalentes nas duas imagens e fazer uma correspondência entre pares de pontos. Para determinar a distância (profundidade) bastará fazer uma triangulação. A utilização dos esboços permite reduzir a possibilidade de ocorrência de correspondências incorrectas pois já lhes foi retirado o "ruído".

Uma das últimas tendências na visão por computador é a utilização de progressos efectuados na pesquisa dos processos fisiológicos que possibilitam a visão nos animais. Já há algum tempo que se sabe que a informação visual não é toda processada no cérebro. O olho não se limita a captar as imagens e a passá-las para o cérebro. A retina, que é composta por várias camadas de neurónios, filtra a informação, rearranja-a e envia pelo nervo óptico somente o que é útil. Ela compara automaticamente os pontos de uma imagem com os pontos circundantes para detectar os contornos dos objectos. Também foi descoberto que os neurónios da retina funcionam igualmente como detectores de características, sendo capazes de automaticamente detectar orientação de linhas, côr e movimento. Tudo isto passa-se sem intervenção do cérebro, por isso diz-se que a detecção destas características constitui a visão primária ou de baixo nível.

Os investigadores estão a chegar à conclusão de que poderão encontrar soluções para os problemas que se lhes deparam se tentarem modelizar as estruturas neuronais dos animais. Assim, pensa-se que o paralelismo massivo obtido com simulações dos processos neuronais terá mais sucesso na simulação dos mecanismos da visão que os algoritmos e programas proporcionados até agora pela Inteligência Artificial.

Os avanços obtidos na pesquisa da visão por computador têm permitido desenvolver aplicações específicas, entre as quais se destacam, por exemplo, sistemas de visão tridimensionais para robots (permitindo uma maior automatização e autonomia nas suas operações), inspecção de materiais perigosos (materiais radioativos), verificação da integridade das conexões microscópicas num circuito integrado, sistemas para orientação de armamento, reconhecimento de alvos de radar, satélites de vigilância, cartografia automatizada, leitores de códigos de barras. Muitas outras aplicações serão desenvolvidas nos próximos anos quando novos avanços na informática permitirem a obtenção de computadores mais potentes e mais acessíveis.

A investigação levada a cabo sobre a visão por computador já é, hoje em dia, muito extensa. No entanto, pode ser dividida em várias temáticas, conforme os problemas a que pretende fornecer soluções. Para facilitar a nossa análise, iremos fazer uma divisão constituída pelas seguintes áreas, que serão abordadas separadamente:

- detecção de características;
- correspondências e obtenção das disparidades;
- movimento e estrutura.

Na realidade, estas áreas não são mais do que as fases que é necessário seguir para simular o processo visual que permite identificar a estrutura tridimensional de objectos e, caso exista, o seu movimento.
Nos sub-capítulos seguintes vamos levar a cabo um levantamento de alguma da pesquisa efectuada nestas áreas de investigação.

3.2. Detecção de Características

Este é o problema da visão por computador que mais tem sido abordado e que se encontra mais detalhadamente documentado. De facto, foi um dos primeiros problemas que se colocaram aos investigadores pois, como já vimos no sub-capítulo anterior, corresponde à visão primária ou de baixo nível e que, nos animais, é levada a cabo no próprio olho. A visão primária, apesar de ser um processo de baixo nível, desempenha um papel muito importante pois é responsável pela obtenção de uma representação intermédia da imagem, a qual será posteriormente interpretada por processos de alto nível.

A escolha da(s) característica(s) a ser(em) utilizada(s) deve ser feita cuidadosamente pois a estratégia a seguir para mais tarde efectuar as correspondências depende dessa escolha. As características mais utilizadas são os cantos e contornos pois são estas que podem ser detectadas com maior facilidade e fiabilidade. Aqueles podem ser facilmente detectados pois, normalmente, onde eles estão situados ocorrem descontinuidades ou mudanças rápidas na intensidade da luminosidade ou nas texturas, quando se passa de um objecto para outro. Estas mudaças são detectadas por um operador local, aplicado sucessivamente a fracções reduzidas da imagem, o qual mede a

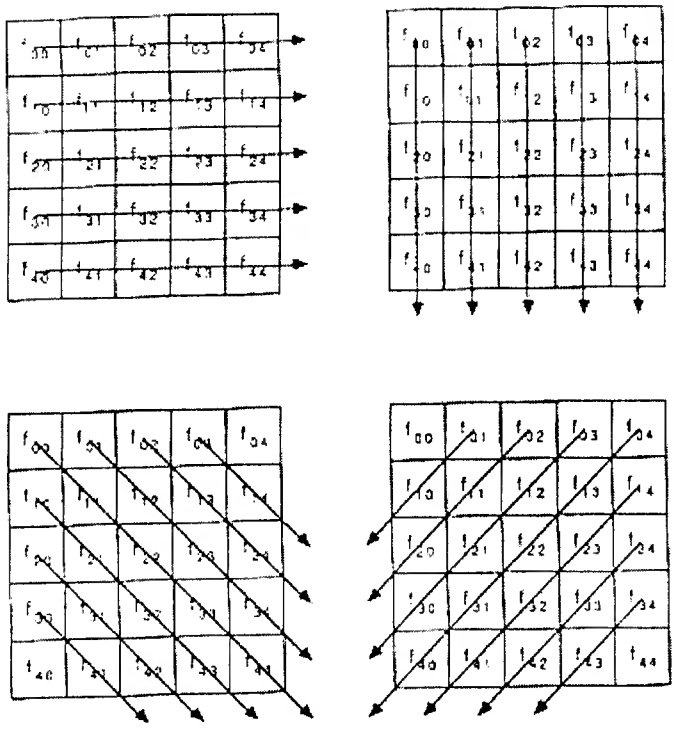
magnitude da mudança e, nalguns casos, a orientação da mudança. Tal como é referido por **Burns, Hanson e Riseman (1986)**, estes operadores apresentam alguns problemas que podem dificultar a detecção, entre os quais se destacam os três principais:

- a dimensão do operador local pode ser reduzida e não abarcar a totalidade da característica que deveria detectar;
- o desvio dos dados presentes na imagem em relação aos modelos assumidos;
- a perda de informação devido ao facto de a digitalização da imagem ser de natureza discreta.

Depois de detectadas as características com interesse numa imagem, passa-se para a imagem seguinte e aplica-se novamente o método de detecção. Numa fase seguinte, procurar-se-á estabelecer-se as correspondências, ou seja, determinar quais as características detectadas na segunda imagem que correspondem, fisicamente, àquelas detectadas na primeira imagem.

Alguns algoritmos utilizam áreas (ou janelas) de tamanho pré-definido para a detecção de características e para posteriormente efectuarem as correspondências. Dentro deste caso, temos, por exemplo, os trabalhos de **Moravec (1977)** e **(1981)**. O algoritmo por ele desenvolvido procura detectar pontos de alto interesse dentro de uma janela, a qual tem uma dimensão definida à partida. Esses pontos de interesse são pontos que, pelo facto de se salientarem mais nas imagens, são mais facilmente detectados. Esses pontos, normalmente, correspondem aos cantos e contornos dos objectos. Para os detectar, utiliza um *operador de interesse*, o qual classifica como pontos de interesse os pontos que apresentem uma grande variância na intensidade da luminosidade em todas as direcções. Para cada *pixel*, o operador calcula os somatórios dos quadrados das diferenças na intensidade da luminosidade dos *pixels* adjacentes nas direcções horizontal, vertical e nas duas diagonais e escolhe o menor desses somatórios como a variância desse ponto. Normalmente, onde existe um ponto de interesse iremos encontrar muitos outros. Para evitar que o número de pontos de interesse seja muito grande e para que se encontrem uniformemente distribuidos por toda a imagem, escolhe-se somente o ponto que apresente a variância mais elevada, dentro de uma janela com uma certa dimensão, como ponto de interesse. A Figura 3.1 mostra-nos como se procederia ao cálculo das variâncias para uma janela de (5x5).

Figura 3.1 - Processo de cálculo das variâncias direccionais



Nota: Extraído de **Nasrabadi e Choo (1992)**

As equações para o cálculo das variâncias para cada direcção serão:

Direcção horizontal (DH): $$\sum_{x=0}^{4}\sum_{y=0}^{3}\left[f(x,y)-f(x,y+1)\right]^2 \qquad (3.1)$$

Direcção vertical (DV): $$\sum_{x=0}^{3}\sum_{y=0}^{4}\left[f(x,y)-f(x+1,y)\right]^2 \qquad (3.2)$$

Direcção diagonal (DD-135°): $$\sum_{x=0}^{3}\sum_{y=0}^{3}\left[f(x,y)-f(x+1,y+1)\right]^2 \qquad (3.3)$$

Direcção diagonal (DD-225°): $$\sum_{x=0}^{3}\sum_{y=1}^{4}\left[f(x,y)-f(x+1,y-1)\right]^2 \qquad (3.4)$$

A função $f(x,y)$ diz respeito à intensidade da luminosidade, numa escala em tons de cinzento, no ponto (x,y). Em cada um dos pontos analisados, o menor valor de DH, DV, DD-135° e DD-225° será seleccionado como a variância nesse ponto.

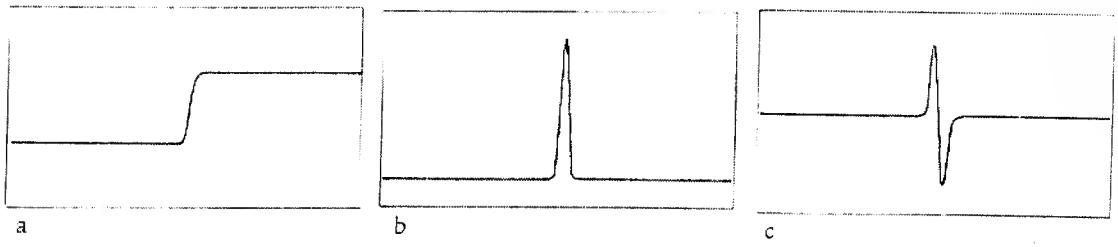
Computacionalmente, o operador de Moravec não é difícil de implementar pois é um algoritmo relativamente simples, quando comparado com outros métodos mais complicados de detecção de cantos.

Dois investigadores em particular influenciaram bastante a pesquisa sobre a visão estereoscópica: David Marr e Tomaso Poggio. A título exemplificativo da pesquisa basilar por eles desenvolvida, dever-se-á consultar os seguintes trabalhos: **Marr e Poggio(1976)**, **Marr e Poggio(1979)**, **Marr (1982)** e **Poggio (1984)**. Parte da pesquisa destes investigadores foi dedicada à detecção de características a utilizar para o estabelecimento de correspondências. Eles sabiam que não poderiam utilizar directamente os valores da intensidade da luminosidade pois estes não são estáveis. Para resolverem esse problema, recorreram a uma propriedade das superfícies dos objectos: nos locais onde existem mudanças físicas nas superfícies (presença de um canto), a imagem dessa superfície apresenta acentuadas variações na intensidada da luminosidade (Figura 3.2-a). Portanto, os cantos são uma característica mais estável do que a luminosidade. Para detectar um canto bastaria detectar mudanças bruscas na luminosidade.

Essas mudanças são detectadas comparando-se a intensidade em dois pontos vizinhos. Matematicamente, resume-se a calcular a primeira derivada. O seu valor representa a taxa de variação da intensidade ao longo de uma certa direcção. Picos na primeira derivada indicam a presença de um canto, pois os picos ocorrem somente onde existam mudanças bruscas na intensidade. Ao longo de uma superfície, a primeira derivada apresentará valores próximos de zero (a intensidade varia pouco). Quando se passa de uma superfície para outra, a primeira derivada apresentará um pico, pois verifica-se uma mudança brusca na intensidade (Figura 3.2-b). Quanto à segunda derivada, indica a taxa de variação da primeira derivada. Ao longo de uma superfície estará também próxima de zero e, ao se aproximar de um canto, aumentará porque a primeira derivada também aumenta. No ponto em que a primeira derivada atinge o seu máximo, a segunda derivada "corta" o zero e começa a assumir valores negativos, pois a

primeira derivada começa a decrescer à medida que a intensidade começa a estabilizar na nova superfície. À medida que a primeira derivada vai decrescendo e começa a se aproximar do zero novamente, também a segunda derivada crescerá até chegar ao zero (Figura 3.2-c).

Figura 3.2 - Derivadas da intensidade da luminosidade



Nota: Extraído de **Poggio (1984)**

Ambas as derivadas enfatizam a presença de um canto. A primeira derivada marca-o com um pico, enquanto a segunda marca-o com um "cruzamento" do zero (mudança de sinal).

No entanto, as derivadas não são suficientes para detectarem a presença de cantos numa imagem real porque muitas vezes as mudanças de intensidade não são suficientemente bruscas ou então estão corrompidas pela presença de "ruído". Para resolverem este problema, aqueles autores fazem um "alisamento" da imagem através do cálculo de uma média dos valores da intensidade de pontos vizinhos. Ambos os procedimentos (diferenciação e "alisamento") são levados a cabo simultaneamente. Para esse fim é utilizado um filtro que incorpora uma função denominada *Laplaceana de Gauss* (LoG). A função de Gauss é a distribuição estatística em forma de sino. Neste contexto, especifica a importância a atribuir à vizinhança de cada *pixel* quando se está a fazer o "alisamento". À medida que a distância aumenta, a importância diminui. A função Laplaceana consiste numa segunda derivada que atribui um peso igual a todas as direcções que se estendem a partir de um ponto. A combinação das duas funções converte a distribuição em forma de sino em algo parecido com um chapéu mexicano. O sino é estreitado e nas pontas desenvolve-se uma reviravolta circular negativa.

Matematicamente, a função LoG é obtida da seguinte forma. Começando pelo operador Laplaceano, a sua representação em coordenadas cartesianas é dada por:

$$\nabla^2 = \frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y}. \tag{3.5}$$

Quanto à função de Gauss, a sua expressão matemática é:

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{3.6}$$

onde σ é um parâmetro de escala. Aplicando o operador Laplaceano à função de Gauss temos:

$$\nabla^2 G(x,y) = \frac{1}{2\pi\sigma^4}\left(\frac{x^2+y^2}{\sigma^2} - 2\right)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{3.7}$$

O processo de filtragem consiste em substituir cada *pixel* por uma média ponderada dos *pixels* vizinhos, sendo os ponderadores fornecidos pela Laplaceana de Gauss. Ao *pixel* é atribuído o ponderador positivo mais elevado e aos *pixels* circundantes ponderadores positivos decrescentes. Depois, segue-se um anel no qual são atribuídos ponderadores negativos. O resultado final da filtragem é um conjunto de valores positivos e negativos, ou seja, uma espécie de segunda derivada da imagem de intensidade. Os "cruzamentos" de zero corresponderão aos locais da imagem original nos quais se verificavam mudanças bruscas na intensidade.

Alguns investigadores preferem detectar não os pontos situados nos cantos dos objectos mas sim linhas rectas presentes nas imagens. As linhas rectas são formadas por um conjunto de pontos colineares e contíguos situados nos cantos, os quais, mais uma vez, traduzem descontinuidades na intensidade da luminosidade. Quanto às linhas curvas, poderão ser representadas como um agregado de vários segmentos de linhas rectas com orientações diferentes.

Depois de detectados os cantos, várias técnicas podem ser utilizadas para a sua agregação em linhas e para a eliminação de toda a informação desnecessária. Essas técnicas incluem transformadas de Hough, detecção de cantos e acompanhamento do contorno, ajustamento de curvas, métodos de grafos, algoritmos de relaxação e técnicas hierárquicas de refinação.

Como exemplo deste tipo de detecção temos o trabalho de **Burns, Hanson e Riseman (1986)**. O modelo por eles desenvolvido teve como ponto de partida dois pontos fracos dos modelos anteriores de extracção de linhas rectas:

- não têm uma visão global da estrutura subjacente da imagem antes de tomarem decisões locais quanto à ocorrência de cantos;
- relegam informação sobre a orientação dos cantos para segundo plano no processamento.

De facto, a maioria desses algoritmos utiliza a magnitude da mudança na intensidade da luminosidade para medir a importância do canto, utilizando a orientação do canto apenas para modular o processo de agrupamento dos cantos "fortes". Os autores atrás referidos consideram que a orientação do canto fornece informação importante acerca do conjunto de *pixels* que participam na variação de intensidade que está subjacente à linha recta, particularmente no que diz respeito à sua extensão.

A orientação do gradiente é definida como a direcção de mudança máxima na escala de tons de cinzento medida numa pequena área à volta de um *pixel* ou, de forma equivalente, a direcção mais acentuada de subida ou descida na superfície da intensidade.

A magnitude do gradiente, medida à volta de um *pixel,* é definida como a mudança máxima na escala de tons de cinzento que ocorre entre esse *pixel e os pixels* adjacentes.

Assim, o algoritmo, baseando-se na orientação do gradiente, tentará detectar regiões na imagem que correspondam a linhas (aquilo que eles denominam "*line-support regions*"). Estas regiões correspondem a vários *pixels* que, agrupados, formarão linhas e constituirão uma superfície de intensidade idêntica. Os *pixels* que constituem estas superfícies têm duas características:

- a magnitude do gradiente (medida numa pequena janela) variará de forma

significativa ao longo da superfície de intensidade, principalmente na direcção ortogonal à linha;

- a orientação do gradiente variará muito pouco ao longo da superfície de intensidade.

O algoritmo baseia-se na consistência da orientação do gradiente para extrair a região de suporte. O acto de isolar estas regiões permite medir com maior precisão outras características da linha, como a sua largura e o seu contraste.

Este método foi originado pela necessidade de extrair linhas rectas de imagens relativamente complexas, particularmente linhas longas e com contraste não muito elevado. O algoritmo segue quatro passos para a extracção de linhas rectas:

1) Agrupar *pixels* em regiões de suporte baseando-se na semelhança da orientação do gradiente

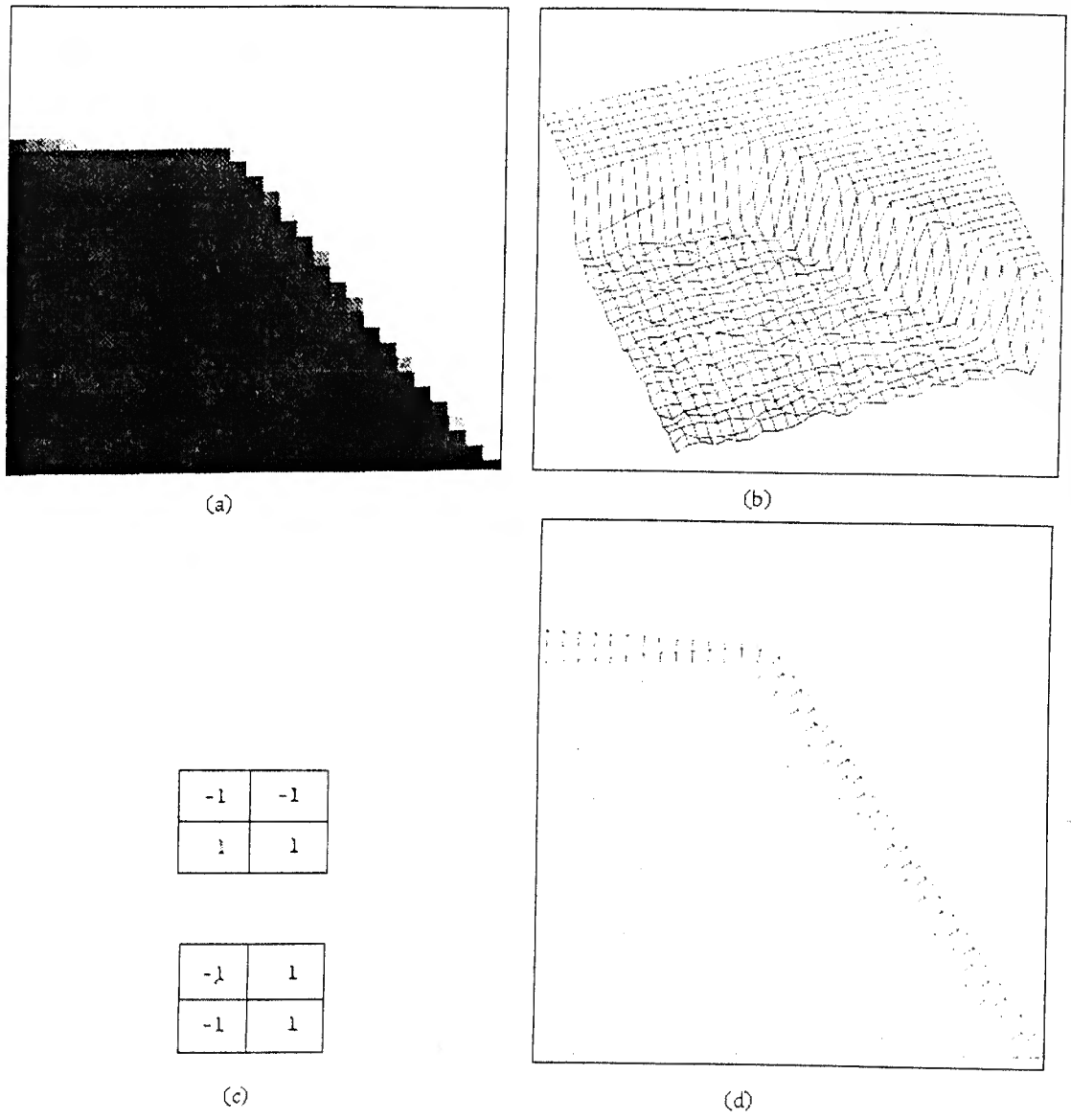
2) Aproximar a superfície de intensidade obtida a uma superfície planar. A superfície planar escolhida é ponderada pela magnitude do gradiente associada aos *pixels* por forma a que as intensidades na parte mais acentuada do canto dominem.

3) Extrair atributos da região de suporte e da superfície planar. Esses atributos incluem a linha ajustada à região, o seu comprimento, largura, contraste, localização, orientação e até que ponto se desvia de uma linha recta.

4) Filtrar linhas para isolar várias ocorrências na imagem, como linhas rectas longas com qualquer contraste, linhas curtas de alto contraste (textura densa), linhas curtas de baixo contraste (textura leve), regiões homogéneas de linhas adjacentes de baixo contraste e linhas em orientações e posições particulares.

A figura 3.3 ilustra o processo de agrupamento e de extracção de linhas rectas.

Figura 3.3 - Processo de agrupamento e extracção de linhas rectas



Nota: Extraído de **Burns, Hanson e Riseman (1986)**

Na Figura 3.3 (a) temos uma janela de (32x32) *pixels* extraida de determinada imagem. A Figura 3.3 (b) mostra a superfície de intensidade obtida para esta sub-imagem e na Figura 3.3 (d) está representada a imagem do gradiente. O comprimento dos vectores codifica a magnitude do gradiente e a sua direcção representa a orientação do gradiente. A magnitude e a orientação do gradiente foram estimadas aplicando à imagem as duas "máscaras" (*masks*) da Figura 3.3 (c). O sinal do gradiente codifica mudanças na intensidade da luminosidade de escuro para claro ou vice-versa. São estes dois operadores de (2x2) que permitem estimar *dI/dx* e *dI/dy*. A orientação do gradiente é calculada pela seguinte expressão:

$$\tan^{-1} G_V(i,j) / G_H(i,j) \tag{3.8}$$

$G_V(i,j)$ e $G_H(i,j)$ são as componentes vertical e horizontal do gradiente obtidas pela aplicação da "máscara" ao *pixel* (i,j).

O passo seguinte consiste na segmentação da imagem do gradiente. Essa segmentação consiste no agrupamento de *pixels* com uma orientação do gradiente semelhante. Para esse efeito, os 360° possíveis de direcção do gradiente são particionados em vários intervalos, por exemplo, de 45° ou de 22,5°. Para cada *pixel*, é analisada a sua orientação do gradiente e determinado o intervalo a que pertence, recebendo um etiqueta. Os *pixels* com a mesma etiqueta e que sejam contíguos formam uma região de suporte para a existência de uma linha recta.

A título exemplificativo da pesquisa desenvolvida em Portugal no domínio da Visão por Computador, refira-se o trabalho de **Oliveira e Ramos (1988)**. Estes investigadores propõem um modelo para a identificação de objectos presentes numa imagem captada de uma cena. Para esse efeito, o seu algoritmo identifica os segmentos de rectas, agrupa-os e classifica as suas junções. De seguida, utiliza um dicionário de junções para etiquetar os segmentos. Segue-se o agrupamento das linhas em regiões e destas em estruturas representativas dos objectos. Por fim, essas estruturas serão comparadas com os modelos flexíveis da base de conhecimento para se obter uma descrição da cena.

3.3. Correspondências e a Obtenção de Disparidades

O processo clássico para a obtenção de visão estereoscópica binocular consiste na captação de imagens por duas câmaras separadas por uma distância fixa, obtendo-se duas imagens simultâneas com ângulos de visão ligeiramente diferentes. A seguir à detecção de características relevantes e determinação das suas posições nas imagens, há que efectuar a correspondência entre pontos homólogos, isto é, entre os pontos que são projecções da mesma identidade física nas imagens analisadas. Depois de efectuadas as correspondências, é analisada a relação espacial entre esses pares de pontos, obtendo-se

as disparidades. A disparidade entre dois pontos consiste na diferença entre as coordenadas de um ponto na primeira imagem e as coordenadas do ponto que lhe é equivalente na segunda imagem. Através de imposição de certas restrições e utilizando as disparidades obtidas, é possível calcular a profundidade da cena, ficando-se assim com as coordenadas que permitem obter uma representação tridimensional da cena.

O problema das correspondências é muito ambíguo pois para um ponto da primeira imagem podem existir muitos outros pontos na segunda imagem que são candidatos plausíveis para a correspondência, podendo ocorrer correspondências erradas. Para se resolver esse problema, impõem-se restrições baseadas nos atributos físicos da cena, limitando-se o número de candidatos para uma correspondência. Assim, ficam definidas as propriedades que uma correspondência correcta deve apresentar. As restrições mais utilizadas são a compatibilidade entre os tipos de características utilizadas para efectuar a correspondência, a unicidade de cada correspondência (cada ponto numa imagem só pode ter um ponto correspondente na outra imagem) e a continuidade das disparidades ao longo da imagem. Esta última restrição resulta da suposição que as superfícies dos objectos são predominantemente suaves. Assim, é esperado que a suavidade na profundidade resulte também na suavidade das disparidades resultantes do processo de correspondências. Para além disso, os contornos na superfície da cena são projectados em cada imagem como curvas contínuas, o que dá origem à restrição da continuidade das figuras, a qual constitui uma restrição de consistência global das correspondências efectuadas localmente.

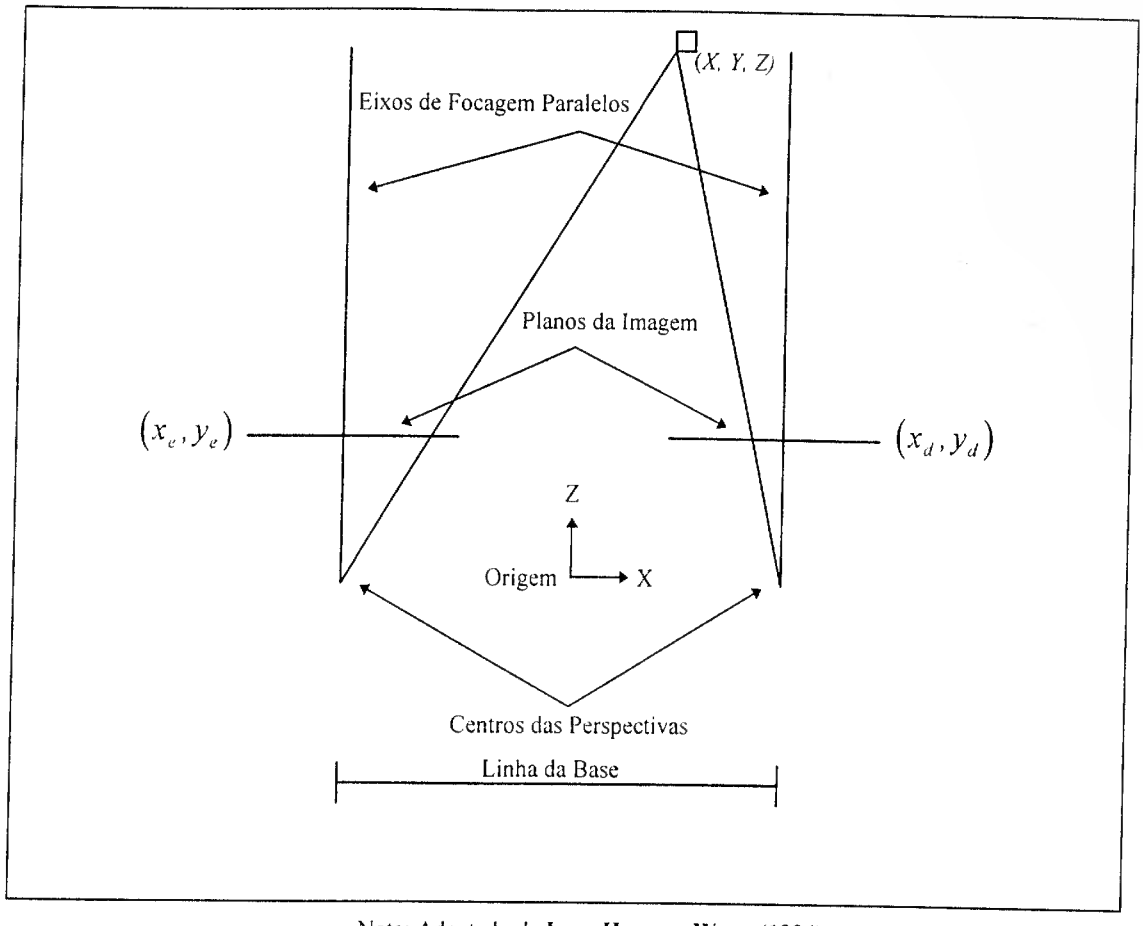
As estratégias utilizadas para o estabelecimento de correspondências podem ser diferenciadas, em termos gerais, pelo tipo de método utilizado e pela geometria do dispositivo de imagem. Quanto ao método, temos o estabelecimento de correspondências utilizando certas características da imagem ou correspondência entre áreas da imagem (*feature-based x area-based*). Quanto à geometria, podemos fazer distinções entre dispositivos com eixos ópticos paralelos entre as câmaras ou eixos não paralelos, ou aínda entre estereoscopia monocular (uma só câmara que se desloca, tirando imagens a intervalos regulares ou irregulares), binocular (duas câmaras) e trinocular (três câmaras). A pesquisa a efectuar para o estabelecimento de correspondências será conduzida de acordo com o tipo de geometria de projecção do dispositivo de imagem. Esses procedimentos de busca serão expressos em termos de restrições epipolares.

As técnicas baseadas em correspondências entre áreas utilizam a correlação entre padrões de intensidade da luminosidade. Primeiro, escolhe-se um ponto de interesse numa imagem e analisa-se a intensidade da luminosidade na sua vizinhança. De seguida, analisam-se áreas equivalentes na outra imagem, escolhendo-se aquela que apresentar a maior correlação com a da primeira imagem. As técnicas basedas em áreas têm a desvantagem de utilizarem directamente o valor da intensidade em cada *pixel,* sendo por isso sensíveis a distorções causadas por mudanças na perspectiva, tal como mudanças na intensidade, contraste e iluminação.

Em contrapartida, técnicas baseadas em características utilizam características simbólicas extraídas das imagens, em vez dos valores da própria intensidade. Assim, estes sistemas são mais resistentes a mudanças de contraste e iluminação. Para além disso, também são mais rápidos que aqueles baseados em áreas, pois permitem simples comparações entre os atributos das características utilizadas para se efectuarem as correspondências. Como já vimos, as características mais utilizadas são os cantos e segmentos lineares.

Quanto à geometria do dispositivo de imagem, existem vários factores que podem ser mudados. Entre eles, temos a orientação dos eixos ópticos das câmaras (paralelos ou não paralelos) e o número de câmaras utilizadas. Um dispositivo de imagem estéreo convencional envolve duas câmaras, com os seus eixos ópticos situados paralelamente um em relação ao outro, e separadas por uma distância horizontal denominada linha da base (*baseline*). Os eixos ópticos das câmaras encontram-se orientados perpendicularmente em relação à linha da base e as suas linhas de digitalização da imagem são paralelas à linha da base. Como a posição relativa das câmaras varia apenas horizontalmente, a posição de pontos correspondentes nas duas imagens também só variará horizontalmente. Portanto, esses pontos situar-se-ão na mesma linha, o que permite reduzir a essa mesma linha a pesquisa para a realização das correspondências. A Figura 3.4 procura demonstrar este tipo de geometria.

Figura 3.4 - Exemplo de uma geometria para visão estereoscópica



Nota: Adaptado de **Lew, Huang e Wong (1994)**

Depois de estabelecida uma correspondência entre dois pontos - (x_e, y_e) na imagem esquerda e (x_d, y_d) na imagem direita - é possível, segundo **Lew, Huang e Wong (1994)**, calcular a posição do ponto correspondente no mundo real (X,Y,Z):

$$X = b(x_e + x_d) / (2(x_e - x_d)) \tag{3.9}$$

$$Y = b(y_e + y_d) / (2(x_e - x_d)) \tag{3.10}$$

$$Z = bf / (x_e - x_d) \tag{3.11}$$

em que b é a distância da base, f é a distância de focagem das câmaras e $(x_e - x_d)$ é a disparidade obtida depois de estabelecida a correspondência entre os dois pontos.

Um problema central na visão por computador é aquilo a que se dá o nome de oclusão. A oclusão dá-se quando objectos, ou parte de objectos, que estão visíveis na primeira imagem analisada não estão presentes na imagem seguinte. Isso pode ocorrer quando esses objectos ficam ocultos por outros objectos que se situam na sua frente ou então quando saem do campo de visão. Tal facto pode originar erros no estabelecimento das correspondências se os algoritmos não estiverem preparados para lidarem com a oclusão. De facto, se um *pixel* que foi detectado na primeira imagem tiver o seu homólogo ocluso na imagem seguinte, o algoritmo poderá estabelecer uma correspondência errada com um outro *pixel* com características semelhantes às do primeiro.

Dhond e Aggarwal (1995) apresentam uma definição formal para a ocorrência de oclusão. Vamos representar por w a largura dos *pixels* de um objecto A responsável por uma oclusão na imagem esquerda de um par de imagens estereoscópicas e por w_N a largura da vizinhança de suporte local utilizada por um algoritmo para retirar a ambiguidade aos vários candidatos à correspondência. Vamos também representar por d_{BG} e d_{FG} as disparidades médias da região oclusa (*background*) e da região que está a provocar a oclusão (*foreground*) na imagem esquerda, respectivamente. Se:

$w < w_N$

e/ou

$w < (d_{FG} - d_{BG})$,

então A é um objecto ocluso.

Estes investigadores também levantam a questão das sombras. Regiões de uma imagem que sejam constituídas por sombras não têm correspondências verdadeiras na outra imagem. Isto resulta do facto que *pixels* situados em zonas com sombras são difíceis de detectar e, para além disso, as sombras não se situarem na mesma zona da cena na imagem seguinte pois podem ter ocorrido mudanças na iluminação ou então pela deslocação da câmara. Portanto, as sombras são um dos factores que contribuem para a existência de "ruído" nas imagens.

Para fazer face a estes problemas, estes investigadores propõem um método denominado Pesquisa Dinâmica de Disparidades que tem por finalidade reduzir o número de correspondências incorrectas em cenas nas quais ocorram oclusões. Esta metodologia define um conjunto de algoritmos baseados em várias escolhas possíveis quanto a configurações de dispositivos de imagem, características a detectar, restrições locais para correspondência e um mecanismo de hierarquia espacial para a imposição das restrições globais de consistência das correspondências. No caso deste último, o algoritmo escolhido baseia-se no algoritmo de Marr, Poggio e Grimson (**Grimson (1981)**), mas os autores afirmam que qualquer algoritmo que se baseie no mecanismo de hierarquia espacial pode ser utilizado.

A metodologia da Pesquisa Dinâmica de Disparidades tem dois princípios básicos. Primeiro, o processo de estabelecimento de correspondências será efectuado separadamente para as superfícies que estão a provocar a oclusão e aquelas que se encontram oclusas, originando grupos separados de disparidades, FG (foreground) e BG background), respectivamente. Isto fará com que elas não interfiram nas vizinhanças de suporte local de *pixels* pertencentes a outras zonas da imagem. Em segundo lugar, fazem variar, dinamicamente, o intervalo de disparidades permitidas, partindo do menor para o maior, o que reduz o número de candidatos à correspondência.

O processo de estabelecimento de correspondências é conduzido separadamente para BF e FG em três fases:

1) Para cada ponto na imagem esquerda, definir uma lista de candidatos à correspondência na imagem direita. Reter apenas aqueles que se situam dentro dos intervalos de disparidade permitidos para os grupos BG e FG.

2) Identificar os pontos que têm apenas uma correspondência dentro dos grupos BG e FG.

3) No caso de correspondências múltiplas, retirar as ambiguidades utilizando a restrição de consistência global do mecanismo de hierárquia espacial. A única diferença em relação aos métodos clássicos é que esta restrição, aqui, é aplicada separadamente em cada um dos grupos de disparidades BG e FG.

Marr e Poggio também abordam o problema das correspondências. A sua formulação computacional baseia-se nos mecanismo biológicos inerentes à visão humana. A teoria por eles desenvolvida propunha que o processamento visual humano resolvia o problema das correspondências em cinco etapas:

1) As imagens esquerda e direita são filtradas por 12 "máscaras" de orientações específicas, sendo cada uma aproximada pela diferença de duas funções de Gauss.

2) Mudanças de sinal (*zero-crossings*) nas imagens filtradas são detectadas através de uma pesquisa em linhas perpendiculares à orientação da "máscara".

3) Para cada tamanho da "máscara", são estabelecidas correspondências entre os cantos que apresentem aproximadamente a mesma orientação e o mesmo sinal.

4) Correspondências obtidas por "máscaras" mais extensas auxiliam o estabelecimento de correspondências por "máscaras" mais pequenas.

5) Os resultados das correspondências são armazenados num *buffer* dinâmico denominado esboço 2.5 D.

Para restringir o problema do estabelecimento de correspondências, Marr e Poggio apresentam duas regras básicas:

- qualquer ponto situado numa dada superfície tem apenas uma localização tridimensional, em qualquer altura, logo a cada ponto numa imagem só pode ser associado um valor de disparidade (*unicidade*);
- a matéria é coesiva e, geralmente, opaca. Logo, os valores da disparidade devem variar suavemente ao longo de uma mesma superfície, excepto quando ocorrem descontinuidades na profundidade provocadas pela mudança de uma superfície para outra (*continuidade*).

Utilizando a técnica por eles desenvolvida e que se encontra descrita no sub-capítulo 3.2, é possível detectar os cantos, os quais vão ser incluídos numa representação da imagem original, a que Marr deu o nome de esboço primário. A partir deste esboço é possível obter informação sobre a posição, direcção, escala e magnitude do gradiente da intensidade, com a qual o algoritmo poderá estabelecer correspondências.

Algum tempo depois, **Grimson (1981)** implementou a teoria computacional desenvolvida por Marr e Poggio, acrescentando mais alguns pormernores que estes não tinham abordado. A estratégia desenvolvida por Grimson para o estabelecimento de correspondências segue um processo iteractivo, passando de resoluções baixas para resoluções mais apuradas. As disparidades encontradas a resoluções baixas serão utilizadas para orientar a pesquisa de correspondências nas resoluções seguintes. Para cada "cruzamento" do zero (que indica a presença de um canto) $P_E(x,y)$ na imagem esquerda, os possíveis candidatos $P'_D(x',y)$ são pesquisados ao longo da linha epipolar na imagem da esquerda de tal forma que:

$$x + d_i - w \leq x' \leq x + d_i + w \tag{3.12}$$

sendo d_i a disparidade estimada e $w = 2\sqrt{2}\sigma$ a largura do filtro baseado na função Laplaceana de Gauss.

Os "cruzamentos" de zero na imagem esquerda e na imagem direita (na mesma linha epipolar) que apresentem o mesmo sinal de contraste e aproximadamente a mesma orientação (dentro de um intervalo de 30°) serão correspondidos. Se só for encontrada uma correspondência dentro da região $\pm w$, então essa correspondência é aceite como desprovida de ambiguidades e a disparidade é registada. Caso seja estabelecida mais do que uma correspondência dentro de $\pm w$, então aquela que apresentar uma disparidade do mesmo tipo (convergente, divergente ou zero) que a disparidade dominante na vizinhança é aceite. Caso contrário, a correspondência permanece ambígua. Depois, os resultados das correspondências são analisados e, caso a percentagem de pontos correspondidos seja menor que 0.7, então todas as correspondências estabelecidas são descartadas. Tal como Marr e Poggio, Grimson impõe uma restrição de continuidade das disparidades referentes à mesma superfície.

Mais tarde, **Grimson (1985)** discute algumas limitações da sua primeira implementação e propõe uma nova, com algumas alterações.

1) *Continuidade figural*: a sua implementação anterior impunha uma continuidade regional nas disparidades para que as correspondências fossem válidas. Grimson observou que esta restrição causava dificuldades na propagação de disparidades nos contornos de objectos sujeitos a oclusão, não ocorrendo problemas noutros locais. Para resolver o problema foi imposta uma restrição de continuidade figural, a qual exige continuidade de disparidades ao longo dos contornos.

2) *Disparidade vertical*: a investigação psico-física indica que o sistema de visão humano recorre a movimentos dos olhos para corrigir desalinhamentos verticais verificados nas imagens. O algoritmo de Marr e Poggio entre em linha de conta com este facto ao pesquisar as linhas epipolares para o estabelecimento de correspondências somente depois de alinhar as duas imagens na direcção vertical. No entanto, distorções locais provocadas por efeitos de perspectiva e "ruído" deterioram a eficácia do estabelecimento de correspondências a níveis de resolução mais elevados. No algoritmo modificado, para um "cruzamento" de zero num ponto $P_E(x,y)$ da imagem esquerda, o ponto correspondente $P'_D(x',y)$ da imagem direita será pesquisado na região

$$\{(x',y'): \quad x+d_i-w \le x' \le x+d_i+w; \quad y-\varepsilon \le y' \le y+\varepsilon \} \tag{3.13}$$

sendo a altura da zona de pesquisa igual a $(2\varepsilon+1)$. Assim, permite-se a existência de uma pequena disparidade vertical (cerca de duas linhas).

Weng, Ahuja e Huang (1992) também abordam, neste seu trabalho, o problema do estabelecimento de correspondências entre duas imagens retiradas da mesma cena. O seu algoritmo utiliza múltiplos atributos presentes em cada imagem para essa finalidade. Esses atributos são representados pelos valores que assumem nas imagens. Estes atributos são submetidos a um processo de tratamento da imagem, pelo qual são "desfocados" para vários níveis de resolução por forma a se obter a informação necessária para o estabelecimento das correspondências.

A utilização de múltiplos atributos permite colocar várias restrições ao problema do estabelecimento de correspondências por forma a que seja perfeitamente

determinável, retirando desta forma as ambiguidades que poderiam conduzir ao estabelecimento de correspondências erradas. Este facto também permite lidar de forma mais eficiente com os problemas causados pelo "ruído" presente nas imagens e com pequenas alterações na intensidade da luminosidade causadas por mudanças na posição de visão, iluminação, reflexos e sombras.

O seu algoritmo também incorpora um mecanismo para lidar com superfícies uniformes e sem texturas, as quais ocorrem frequentemente em imagens do mundo real. Outro problema para o qual o algoritmo apresenta uma solução é a preservação de descontinuidades das disparidades. Por vezes, certas restrições podem conduzir a um alisamento das superfícies dos objectos, perdendo certas descontinuidades que possam existir.

A primeira questão que se coloca é que atributos deverão ser utilizados. Diz-se que um atributo é invariante em relação ao movimento (deslocação da câmara ou dos objectos presentes na imagem) se o seu valor não se alterar de imagem para imagem seja qual for o movimento verificado. A utilização destes atributos como critério para o estabelecimento de correspondências seria o ideal. No entanto, na generalidade das situações, é impossível encontrar um atributo com estas características. Assim, os autores recorrem àquilo a que denominam atributos insensíveis ao movimento, ou seja, atributos que apresentam apenas pequenas variações provocadas pelo movimento. Em certas condições, a intensidade da luminosidade é insensível ao movimento. Assim, regiões das imagens que sejam correspondentes apresentarão valores semelhantes de intensidade. A este princípio, os autores dão o nome de *critério de semelhança da intensidade*.

No entanto, se o estabelecimento de correspondências se basear apenas na intensidade da luminosidade, um ponto poderá ser correspondido com qualquer outro ponto que apresente uma intensidade com um valor semelhante. Normalmente, na mesma imagem, existem vários pontos nesta condições que se apresentam como candidatos para apenas uma única correspondência. Isso é especialmente verdade no caso de pontos que se situem na mesma superfície de determinado objecto, pelo que, normalmente, existem vários deles na mesma vizinhança com valores semelhantes. Portanto, este critério, isolado, é manifestamente insuficiente.

O que se passa na visão humana é que as correspondências não são efectuadas somente com base na intensidade de pontos tomados individualmente. As

correspondências são estabelecidas, ao longo de certo período, com base na estrutura das imagens, ou seja, com base nas relações espaciais entre os pontos da imagem. Portanto, informação estrutural de alto nível (forma da região e relações espaciais entre regiões) é útil para o estabelecimento de correspondências.

O problema que se coloca é que este tipo de informação é muito instável quando se verifica a ocorrência de movimento e de oclusões. Assim, devem ser procurados atributos estruturais de baixo nível que sejam insensíveis ao movimento e que impliquem a utilização de apenas uma pequena vizinhança em torno de um ponto por forma a que a oclusão não provoque erros significativos numa grande área. Os autores propõem a utilização dos cantos como informação estrutural. Como já vimos, os cantos estão associados a acentuadas alterações de intensidade de luminosidade, correspondentes ao fim de uma superfície e ao início de outra. Os cantos são insensíveis ao movimento uma vez que, geralmente, uma alteração acentuada na intensidade presente numa cena permanecerá acentuada mesmo depois de se ter verificado um movimento moderado. Os autores dão o nome de *critério de semelhança de cantos* ao critério que diz que um canto deve ser correspondido a outro que apresente uma medida semelhante. A medida utilizada é a magnitude do gradiente da intensidade.

No entanto, os autores consideram que a semelhança na intensidade e a semelhança nos cantos aínda não são suficientes para o estabelecimento de correspondências correctas. Como atributo adicional, propõem a utilização da forma dos contornos, ou seja, os ângulos por eles formados, para a identificação de vértices presentes nos objectos. Um ponto situado num vértice apresentará uma medida com um valor absoluto superior a outros pontos situados ao longo do contorno do objecto. Além disso, o sinal de um vértice deverá permitir distinguir um vértice de um rectângulo branco sobre um fundo preto do vértice de um rectângulo preto sobre um fundo branco. Os autores dão o nome de *critério de semelhança de ângulos* ao critério que diz que um ponto deve ser correspondido com outro que apresente uma medida de ângulo semelhante à sua. Essa medida tem por base mudanças na direcção do gradiente em dois pontos vizinhos, ponderada pela magnitude do gradiente no ponto. Esses dois pontos estão situados sobre um circulo centrado no ponto. O raio do circulo é determinado pelo nível de resolução utilizado. Os dois ponto são escolhidos por forma a que as derivadas direccionais ao longo do circulo atinjam os seus valores mínimo e máximo.

Assim, o algoritmo por eles apresentado utiliza a intensidade, cantos e ângulos, em conjunto, como atributos para o estabelecimento de correspondências. No entanto, a metodologia por eles utilizada permite a introdução de outros atributos como, por exemplo, a côr.

Para o tratamento de imagens do mundo real, o algoritmo tem que ser capaz de tratar regiões de intensidade uniforme. Normalmente, estas regiões resultam de uma mesma superfície contínua. Este facto permite definir o *critério da suavidade intraregional*. Pontos com a mesma intensidade, situados na mesma região, em princípio devem pertencer à mesma superfície. Este critério não deve ser imposto entre regiões diferentes, pois isso conduziria a correspondências erradas.

Outro problema que este algoritmo procura solucionar é a ocorrência de oclusões. Se as regiões oclusas não forem detectadas, podem ser correspondidas de forma errada com outra regiões. Para evitar isso, é necessário identificar as regiões sujeitas a oclusão numa ou na outra imagem. Com essa finalidade, os autores definem dois mapas de oclusão: o mapa 1, no qual estão incluídas as regiões da primeira imagem que não são visíveis na segunda (ou porque estão tapadas por outros objectos ou porque sairam do campo de visão da câmara) e um mapa 2 para as regiões da segunda imagem que não estão visíveis na primeira. Depois de obtido o mapa 1, procura-se estabelecer correspondências para todas as regiões da primeira imagem, excepto para aquelas que fazem parte do mapa, ou, de forma inversa, para a segunda imagem.

Uma das características deste algoritmo é que permite a existências de disparidades elevadas entre as duas imagens. Normalmente, outros algoritmos só permitem pequenas disparidades como forma de reduzir as áreas de pesquisa para o estabelecimento de correspondências. Nestes casos, o movimento verificado entre o espaço de tempo de recolha das duas imagens não pode ser elevado para evitar que as regiões a corresponder saiam da área de pesquisa. Para permitir a utilização de disparidades elevadas é necessário que se conheça a localização aproximada das correspondências pois, caso contrário, poderão ser estabelecidas múltiplas correspondências.

Uma possível solução é a "desfocagem" como forma de filtrar a imagem. Esta solução apresenta a desvantagem de numa imagem "manchada" restarem poucas características, sendo as suas localizações imprecisas. Por isso, em vez de primeiro "desfocarem" a imagem e depois tirarem as medidas de cantos e de ângulos, os autores

extraem da imagem inicial imagens composta pelos cantos e ângulos (imagens dos atributos). Como a medida dos ângulos tem um sinal, é possível "desfocar" aqueles que tenham valores positivos e negativos próximos por forma a ficarem com valores muito próximos de zero. Assim, só os ângulos significativos serão aproveitados. Por outro lado, é possível agrupar os ângulos com sinal positivo numa imagem e os ângulo com sinal negativo noutra. Teremos, portanto, três tipos de imagens dos atributos: imagens com os cantos, imagens com ângulos positivos e imagens com ângulos negativos. São estas imagens dos atributos que vão ser sujeitadas ao processo de "desfocagem". Depois de "desfocadas", estas imagens, ao contrário da imagem original composta por valores de intensidade, não perdem informação acerca das texturas. Esta informação irá ser utilizada para o estabelecimento das correspondências a um nível de resolução "grosseira", formada por uma grelha. Em níveis seguintes, a resolução das imagens vai aumentando. De nível para nível, cada posição na grelha do nível anterior será dividida em quatro novas posições, e assim sucessivamente, até se chegar à resolução da imagem original. À medida que o nível de resolução vai melhorando, o processo de estabelecimento de correspondências vai ficando mais preciso.

Lew, Huang e Wong (1994) apresentam um algoritmo potente para a selecção de características a utilizar para o estabelecimento de correspondências. As correspondências serão efectuadas utilizando pontos e certos atributos encontrados numa janela situada em torno dos pontos. O objectivo é escolher um conjunto óptimo de características que definam um ponto e que o permitam identificar de forma inequívoca para o estabelecimento de uma correspondência. As características que podem ser utilizadas são as que se encontram especificadas na Tabela 3.1.

Tabela 3.1 - Conjunto de características utilizadas para identificar um ponto

Característica	Descrição	Fórmula Matemática
Intensidade	A intensidade em escala de cinzento no *pixel* (x, y)	$I = h(x, y)$ sendo h a imagem discreta e I a intensidade
Gradiente de X	A primeira derivada de intensidade em ordem a x	$G_x = \frac{\partial I}{\partial x}$
Gradiente de Y	A primeira derivada de intensidade em ordem a y	$G_y = \frac{\partial I}{\partial y}$
Magnitude do gradiente	A magnitude dos gradientes de x e de y	$G_M = \left(G_x^2 + G_y^2\right)^{\frac{1}{2}}$
Orientação do Gradiente	O ângulo formado pelos gradientes de x e de y	$G_o = \tan^{-1}\left(\frac{G_y}{G_x}\right)$
Laplaceana	A Laplaceana da intensidade	$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$
Curvatura	A curvatura no *pixel* s é a curva ao longo de G_o	$C_o = \frac{\partial G_o}{\partial s}$

Nota: Adaptado de **Lew, Huang e Wong (1994)**

O processo para a escolha de um subgrupo de características é composto por três fases:

1) É encontrado um sub-conjunto inicial de características através da aplicação de um algoritmo para esse fim. O *input* desse algoritmo é constituído pelos valores das várias características no ponto em questão e o seu *output* será um sub-conjunto inicial de características a utilizar.

2) Nesta fase, o sub-conjunto de características obtido na fase anterior será refinado através de um processo que permita encontrar o conjunto que maximize a precisão do processo de correspondências, ou seja, que permita identificar o ponto em questão como sendo único.

3) Na última fase, o conjunto de características obtido na fase anterior será utilizado para o estabelecimento de uma correspondência. Caso o conjunto de características seja ambíguo, isto é, caso exista a possibilidade de confundir o ponto em questão com outros pontos e produzir uma correspondência errada, será necessário retirar essa ambiguídade utilizando conhecimento *a priori* sobre superfícies.

Para além do processamento de imagens por computador através de *software* apropriado, alguns investigadores têm desenvolvido *hardware* específico para a simulação da visão. Métodos tradicionais baseados em *hardware* não são adequados para solucionar o problema das correspondências na visão estereoscópica em tempo real pois estes algoritmos exigem uma grande capacidade computacional. Recentemente, alguns investigadores têm desenvolvido *hardware* baseado em VLSI (*very large scale integration*) para ultrapassar este problema. Esta nova tecnologia tem uma maior capacidade computacional que os processadores tradicionais pois implementa um método de processamento em paralelo.

Como exemplo nesta área de investigação, temos o algoritmo implementado em *hardware* desenvolvido por **Erten e Goodman (1996)**. Neste trabalho, apresentam um esquema para um *chip* basedo em VLSI destinado ao estabelecimento de correspondências e cálculo de disparidades.

O algoritmo de Erten e Goodman procura estabelecer correspondências entre pares de imagens captadas por um dispositivo para visão estereoscópica. Uma região da primeira imagem é seleccionada, sendo depois comparada com regiões da segunda imagem que sejam candidatas ao estabelecimento da correspondência, seleccionando-se apenas uma para esse efeito. Antes desse processamento, as imagens são filtradas por um filtro de tipo exponencial para reduzir os efeitos nocivos do "ruído". Os valores filtrados dos *pixels* são utilizados para comparar vizinhanças em cada imagem. Em cada valor possível para a disparidade, é efectuada uma comparação entre as duas vizinhanças filtradas. O algoritmo pode ser descrito da seguinte forma:

1) Seleccionar uma região X na imagem esquerda. X pode ser interpretado como um vector com N elementos, $x_1,\ldots,x_N$.

2) Comparar X com as K regiões candidatas, ou sub-imagens $Y_1,...,Y_i,...,Y_K$ da imagem direita. Cada uma desta regiões Y tem o mesmo tamanho que X.

3) Seleccionar uma região Y como par para X, baseando-se nos valor(es) de uma medida de semelhança.

4) Obter a confiança computacional da correspondência estabelecida, ou seja, quantificar o grau de certeza de se tratar de facto de uma correspondência correcta.

5) Repetir para todas as regiões X da imagem esquerda.

3.4. Movimento e Estrutura

Neste sub-capítulo vamos tratar do problema da obtenção da estrutura tridimensional de uma cena a partir da análise do movimento que ocorre nessa mesma cena. Esse movimento pode ser o movimento realizado pelos objectos presentes na cena, mantendo-se a câmara fixa, ou então pode ser devido a uma movimentação da câmara, podendo também existir movimento dos próprios objectos. As sucessivas imagens obtidas ao longo de um certo período de tempo serão analisadas para a obtenção da estrutura tridimensional da cena. Estabelece-se, portanto, o princípio de que é possível obter essa estrutura somente a partir da análise de imagens bidimensionais de uma cena. Este princípio baseia-se na análise do próprio sistema visual humano. De facto, o sistema visual é capaz de fazer inferências acerca da estrutura tridimensional de uma cena analisando apenas as sua projecções numa série de imagens bidimensionais obtidas em perspectivas diferentes e ao longo de um certo período de tempo. Quanto maior o número de imagens a que o observador for exposto, mais correcta será a inferência efectuada.

Esta fase corresponde ao último passo para a obtenção da estrutura tridimensional. De facto, antes de mais, será necessário detectar pontos característicos em cada imagem (1ª fase) e depois estabelecer as correspondências entre os pontos

característicos detectados nas várias imagens (2ª fase). Somente depois se poderá partir para a obtenção da estrutura a partir da análise do movimento presente na cena.

O problema da inferência gradual acerca da estrutura tridimensional de uma cena a partir da análise de uma sequência de imagens é designado pelos especialistas como o problema do movimento de longo alcance (*long-range motion*). As estratégias desenvolvidas para solucionar este problema podem ser agrupadas em dois grupos:

No primeiro grupo, temos estratégias que utilizam directamente o conteúdo da imagem para determinar a estrutura tridimensional e obter uma representação dos objectos, a qual deverá respeitar uma certa restrição física ou ambiental, como por exemplo, a rígidez dos objectos ou a suavidade das suas superfícies. Rígidez de um objecto refere-se à estabilidade na sua forma, a qual não se deve alterar ao longo da sequência de imagens sujeitas a análise. A restrição referente à suavidade resulta da observação de que, normalmente, a superfície de uma objecto é contínua e suave. As descontínuidades ocorrem normalmente quando termina a superfície de um objecto.

No segundo grupo, temos estratégias que utilizam informação presente nas imagens bidimensionais para extrair os parâmetros referentes ao movimento relativo dos objectos (movimentos de translacção e rotação, por exemplo). O objectivo é a obtenção de informação acerca do movimento tridimensional dos objectos.

Para exemplificar o primeiro tipo de estratégias vamos analisar um algoritmo desenvolvido por **Ullman (1984)** denominado esquema da rígidez incremental. Antes de ser aplicado, exige que previamente já tenham sido estabelecidas correspondências entre os pontos característicos de duas imagens consecutivas.

O algoritmo desenvolvido por Ullman baseia-se no princípio que o sistema visual humano é capaz de extrair informação sobre a estrutura tridimensional a partir de transformações bidimensionais da imagem. Para tal basta que exista movimento na cena (da câmara e/ou dos objectos). No processo de captação de imagens verifica-se uma perda de informação acerca da profundidade da cena, uma vez que as imagens captadas são bidimensionais, enquanto que a cena real é a três dimensões. O que se passa é que ocorre uma projecção da cena tridimensional para o plano bidimensional da imagem captada. Na visão estereoscópica existem dois tipos de projecção: a projecção ortográfica e a projecção central. No caso da projecção ortográfica, os eixos de focagem do

dispositivo de captação da imagem são paralelos entre si e perpendiculares ao plano da imagem. Quanto à projecção central, os eixos de focagem cruzam-se num ponto comum, como é o caso dos olhos no sistema visual humano. O tipo de projecção utilizado influenciará a forma como será feita a inferência sobre a estrutura tridimensional.

Ullman, bem como outros investigadores anteriores a ele, tinha conhecimento de que, ao tentar recuperar a estrutura tridimensional a partir das transformações na imagem, se depararia com o problema da ambíguidade que lhe é inerente. Para solucionar este problema seria necessário impôr algumas restrições, caso contrário as transformações, por si só, seriam insuficientes para determinar a estrutura de forma correcta. Desde os primeiros estudos empíricos sobre este problema foi sugerido que a rígidez dos objectos poderia ser importante na percepção da estrutura a partir do movimento. Posteriormente, estudos computacionais demonstraram que a rígidez constitui uma restrição suficientemente potente para impôr unicidade na interpretação tridimensional das transformações na imagem. Portanto, impondo somente a restrição referente à rigidez dos objectos, é possível recuperar a estrutura tridimensional de uma cena recorrendo apenas a informação sobre o movimento.

Através da análise dos mecanismos visuais humanos, Ullman sugeriu que para que qualquer processo de recuperação da estrutura tridimensional a partir do movimento apresente uma performance comparável ao sistema visual humano deverá satisfazer os seguintes requisitos:

1) Em cada momento deverá existir uma estimativa da estrutura tridimensional dos objectos. Inicialmente, este modelo interno da estrutura visível pode ser grosseiro e pouco exacto, podendo ser influênciado por fontes estáticas de informação tridimensional.

2) O processo de recuperação da estrutura deverá dar preferência a transformações rígidas.

3) O esquema de recuperação da estrutura deverá ser capaz de tolerar desvios da rígidez. Esses desvios podem ocorrer devido a distorções provocadas por "ruído". Se o algoritmo tolerar esses desvios, terá uma certa imunidade ao "ruído" muitas vezes presente nas imagens.

4) Deverá ser capaz de integrar informação referente a um período de visualização extenso. Tal como no sistema visual humano, quanto maior for o número de imagens disponíveis mais correcta será a estrutura extraída. Normalmente, para períodos curtos, as estruturas extraídas são mais achatadas do que as estruturas reais.

5) Eventualmente, deverá ser capaz de recuperar a estrutura tridimensional correcta, ou então uma aproximação aceitável.

Ullman propõe o esquema da rígidez incremental, o qual, segundo ele, satisfaz a maior parte destes requisitos de uma forma natural. O esquema por ele proposto assume que em qualquer momento existe um modelo interno do objecto visionado. *M(t)* representa o modelo interno no momento *t*. À medida que o objecto se move, a sua projecção vai mudando. Se *M(t)* não for um modelo preciso do objecto no momento *t*, então nenhuma transformação rígida de *M(t)* seria suficiente para dar resposta à tranformação verificada na imagem. O passo crucial do esquema consiste em aplicar então uma modificação mínima (não rígida) que seja suficiente para representar a transformação observada. Portanto, o modelo interno, apesar de tentar respeitar o mais possível a restrição da rígidez, vai sofrendo alterações não rígidas (o mais pequenas possíveis) até que se aproxime da estrutura correcta. Noutras palavras, o modelo interno resiste a mudanças tanto quanto for possível, e, consequentemente, torna-se tão rígido quanto for possível.

O modelo interno *M(t)* consiste num conjunto de coordenadas tridimensionais (X_i, Y_i, Z_i). Assumindo a utilização da projecção ortográfica no plano *X* - *Z* da imagem, (X_i, Z_i) são as coordenadas do i-ésimo ponto na imagem e Y_i é a sua profundidade, estimada pelo modelo corrente. Perante a falta de informação tridimensional acerca do objecto, o modelo inicial *M(t)* no momento $t = 0$ apresenta-se completamente achatado, ou seja, $Y_i = 0$ para i = 1 ... *n*, onde *n* é o número de pontos considerados na computação.

De seguida, é considerada uma nova imagem correspondente a um momento posterior t', consistindo o problema em actualizar *M(t)* por forma a reflectir a nova imagem e mantendo a transformação de *M(t)* para $M(t')$ tão rígida quanto possível. A nova imagem é representada por um conjunto de coordenadas bidimensionais (x_i, z_i). Os novos valores da profundidade y_i aínda não foram determinados. No entanto,

assume-se que as correspondências entre pontos das duas imagens sucessivas são conhecidas. Depois de se estimar os valores de y_i, o conjunto de coordenadas (x_i, y_i, z_i) representa a estrutura estimada no momento t', identificada pela notação $S(t')$. É utilizada a convenção de que os parâmetro respeitantes a $M(t)$ são representados por letras maiúsculas (X, Y, Z) e aqueles que se referem a $S(t)$ serão representados por letra minúsculas (x, y, z).

A transformação o mais rígida possível do modelo interno $M(t)$ será determinada da seguinte forma. Representando a distância entre os pontos i e j, pertencentes a $M(t)$, por L_{ij}, teremos:

$$L_{ij}^2 = (X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2. \tag{3.14}$$

De forma semelhante, l_{ij} representa a distância interna na estrutura estimada entre os pontos i e j no momento t'. Teremos então:

$$l_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2. \tag{3.15}$$

Uma transformação rígida implica que $L_{ij} = l_{ij}$ para todos os pontos i e j, ou seja, todas as distâncias internas do objecto mantém-se inalteradas. Para que a transformação seja tão rígida quanto possível, os valores das profundidades y_i deverão ser escolhidos por forma a que os valores de L_{ij} e de l_{ij} estejam o mais próximo possível um do outro. Se $D(L_{ij}, l_{ij})$ for uma medida da diferença entre L_{ij} e l_{ij}, então o problema da determinação da transformação o mais rígida possível do modelo poderá ser formulado como a determinação dos valores de y_i por forma a minimizar o desvio global da rígidez, o qual será dado por

$$\sum_{i,j} D(L_{ij}, l_{ij}) \qquad (i = 1, \ldots, n-1; j = i+1, \ldots, n). \tag{3.16}$$

Segundo Ullman, este desvio da rígidez pode ser utilizado também como uma medida de "confiança": quanto menor for o desvio maior será a certeza de que o modelo interno traduz de forma correcta a estrutura tridimensional do objecto.

A questão que se coloca agora é a escolha da medida *D*. Ela deverá ser definida por forma a que as contribuições dos pontos mais próximos tenham um peso superior aquelas de pontos mais distantes. Isto advém do facto de que os vizinhos mais próximos de determinado ponto terem uma maior probabilidade de pertencerem ao mesmo objecto do que pontos mais distantes. Ullman apresenta um exemplo de uma medida de distância com essa característica, utilizando a seguinte expressão:

$$D(L_{ij}, l_{ij}) = \frac{(L_{ij} - l_{ij})^2}{L_{ij}^3} \tag{3.17}$$

Depois de os valores de y_i terem sido determinados com a utilização do critério de minimização, (x_i, y_i, z_i) passará a ser o novo modelo $M(t')$. De seguida, passa-se para a imagem seguinte e o processo volta-se a repetir. Quanto maior for o número de imagens utilizadas, mais correcta será a estrutura tridimensional extraída.

No segundo tipo de estratégias temos, como já foi referido, métodos cujo objectivo é extrair, a partir de imagens bidimensionais, informação acerca do movimento tridimensional de objectos presentes numa cena. Os processos desenvolvidos envolvem a estimação da natureza e parâmetros do movimento tridimensional, possibilitando a previsão das posições futuras de objectos em movimento.

Este tipo de pesquisa tem várias aplicações, das quais temos os seguintes exemplos:

- pode ser utilizada para a orientação contínua de câmaras que deverão captar autonomamente o movimento de determinado objecto;
- o movimento de um braço de um robot ou de um veículo poderá ter que ser estimado e previsto para o planeamento de trajectos seguros;
- a recuperação e reparação de satélites no espaço exigem que o seu movimento seja conhecido para que sejam abordados pelo veículo que irá levar a cabo essas operações;

- monitorização de células e do coração;
- rastreamento de núvens e de perturbações atmosféricas;
- determinação de correntes marítimas;
- monitorização de tráfego e acompanhamento automático de veículos;
- detecção de alvos militares e seu acompanhamento.

O problema que se coloca é caracterizar quantitativamente, em termos gerais, o movimento tridimensional de objectos a partir da análise das suas imagens bidimensionais. A tentativa de generalizar para todos os casos implica certas dificuldades que advêm da falta de dados acerca da estrutura dos objectos tal como do tipo de movimento a que estão a ser sujeitos. Por exemplo, pode até nem se saber se um objecto está sujeito a um movimento de translacção e/ou de rotação. Se forem aplicadas alguma restrições, o problema ficará mais fácil de resolver mas a solução obtida poderá ter uma utilidade restrita. Para simplificarem o problema, alguns investigadores impuseram restrições quanto ao movimento permitido e quanto à estrutura dos objectos, o que, muitas vezes, tornou a solução inaplicável a situações reais.

Weng, Huang e Ahuja desenvolveram um algoritmo cujo objectivo é obter informação acerca do movimento com o mínimo possível de conhecimento *a priori* (ver **Weng, Huang e Ahuja (1987)** e **Weng, Huang e Ahuja (1989)**). O movimento de um objecto é determinado pela dinâmica que lhe está subjacente. Através da análise de uma sequência de imagens e aplicando um modelo dinâmico generalista, é possível conhecer o movimento presente nessa sequência. Para além disso, baseando-se nos parâmetros do movimento extraídos, será possível fazer extrapolações e interpolações para prever o movimento e para recuperar partes desse movimento que não estejam presentes na sequência.

Em geral, não são conhecidas as forças que actuam sobre o objecto e a resposta estrutural do objecto a essas forças. Portanto, é necessário impor uma restrição ao movimento do objecto para que o problema da inferência acerca do movimento tridimensional tenha uma solução. Weng, Huang e Ahuja fazem a observação de que, geralmente, os objectos apresentam um movimento suave, ou seja, os parâmetros do movimento entre imagens consecutivas estão correlacionados. Os autores afirmam que, tendo por base esta suposição e dada uma sequência de imagens de um objecto rígido em movimento, é possível determinar qual o tipo de movimento a que esse objecto está

sujeito. Para este fim, desenvolveram um modelo a que deram o nome de *Modelo de Ímpeto Angular Localmente Constante*. Este modelo assume a conservação de ímpeto angular no curto prazo e uma curva polinomial para a trajectória do centro de rotação. Esta restrição traduz exactamente aquilo que se entende por suavidade do movimento. No entanto, os autores permitem que, no longo prazo, o ímpeto angular e, consequentemente, as características do movimento se alterem. Assim, o movimento do objecto não é sujeito a uma restrição através da imposição de um modelo global de dinâmica permitida.

Através da utilização deste modelo é possível responder às seguintes questões:

- se o objecto está a rodar para a frente ou para trás;
- como o centro de rotação do objecto (o qual pode ser um ponto invisível) se move no espaço;
- qual será provavelmente o movimento futuro do objecto;
- em que posição nos fotogramas ou no espaço tridimensional estará localizado determinado ponto nos instantes seguintes;
- onde o objecto estaria caso não esteja presente numa sub-sequência da imagem;
- qual seria o movimento anterior à actual sequência.

O facto de ser possível prever a posição futura de pontos característicos permite que somente uma pequena vizinhança da posição prevista seja pesquisada para o estabelecimento de correspondências entre imagens sucessivas. Para além disso, a imposição da restrição de suavidade local no movimento é útil para evitar erros causados pelo "ruído". Quaisquer desvios em relação ao movimento considerado normal já se sabe que serão causados pelos efeitos do "ruído", pelo que se torna mais fácil contrariar esses efeitos. A utilização de um número grande de imagens na sequência também permite combater o "ruído".

Weng, Huang e Ahuja definem o problema da estimação do movimento a partir da análise de duas imagens diferentes obtidas a partir da mesma cena da seguinte forma: dadas imagens de um objecto em movimento captadas em dois momentos diferentes, o problema consiste em estimar a transformação da posição tridimensional do objecto entre esses dois instantes. Para aplicarem o seu modelo, os autores assumem que existe um único objecto em movimento, as correspondências de pontos entre imagens são

dadas e o movimento não apresenta descontinuidades como aquelas causadas por colisões.

A aplicação do modelo permite calcular os parâmetros que definem o movimento de um objecto: a sua rotação e a sua translacção. A rotação será representada pela matriz R e a translacção pelo vector T. Considerando um ponto P pertencente ao objecto que seja visível em dois momentos diferentes e representando por $\boldsymbol{x} = (x, y, z)$ as coordenadas de P no momento t_1 e por $\boldsymbol{x}' = (x', y', z')$ as coordenadas de P no momento t_2, a relação entre as coordenadas de P nos dois momentos será dada por

$$\boldsymbol{x}' = \boldsymbol{R}\,\boldsymbol{x} + \boldsymbol{T} \tag{3.18}$$

Os valores da variável z, relativos à profundidade, podem ser determinados mas só em termos relativos. Teriamos então

$$\boldsymbol{Z}_i = \left(\frac{z_i'}{\|\boldsymbol{T}\|}, \frac{z_i}{\|\boldsymbol{T}\|} \right). \tag{3.19}$$

Esta última expressão permite a obtenção da estrutura tridimensional a partir da análise explícita dos parâmetros do movimento.

4. Aplicação das Redes Neuronais à Resolução do Problema da Obtenção da Estrutura Tridimensional dos Objectos

Este capítulo é dedicado à aplicação das redes neuronais à resolução do problema em estudo. Em primeiro lugar, será feito um enquadramento do problema, seguindo-se uma ligeira abordagem à teoria das redes neuronais. Por fim, neste capítulo, serão apresentados os modelos baseados em redes neuronais para a resolução do problema da obtenção da estrutura tridimensional dos objectos.

4.1. Enquadramento do Problema

O problema central que se coloca no âmbito deste trabalho é a obtenção da estrutura tridimensional de objectos a partir da análise de representações bidimensionais desses mesmos objectos. O processo de captação de imagens implica a perda de informação relativa à profundidade da cena captada. A posição dos objectos no mundo real pode ser representada por um sistema de coordenadas, (X - Y - Z). Se Z representar os valores relativos à profundidade, esta coordenada será perdida no processo de captação de imagem, uma vez que os objectos são projectados no plano da imagem captada, o qual é representado por um sistema de coordenadas bidimensional, (X - Y). O problema a resolver consiste em conseguir recuperar os valores da coordenada Z dispondo apenas de informação relativa à posição dos objectos no plano (X - Y). Isso só poderá ser possível se dispormos de pelo menos duas imagens captadas em ângulos diferentes (a visão estereoscópica exige pelo menos duas vistas da mesma cena captadas em ângulos diferentes para inferência acerca da tridimensionalidade). Quanto maior o número de imagens maior será a fiabilidade das coordenadas obtidas.

A implementação deste sistema exige que sejam seguidas várias etapas. Em primeiro lugar, será necessário processar cada imagem para a detecção de pontos característicos (neste caso, as características utilizadas serão os cantos). Depois de detectados, serão estabelecidas as correspondências entre os pontos extraídos em cada

imagem. Finalmente, e tendo por base as correspondências estabelecidas na etapa anterior, será efectuada uma inferência sobre a estrutura tridimensional dos objectos.

Para a resolução do problema em questão, este trabalho propõe a utilização de redes neuronais. As abordagens clássicas a este tipo de problema baseiam-se em algoritmos sequênciais, os quais poderão não ser a solução mais adequada e mais rápida para a resolução do problema da visão por computador. De facto, basta comparar o desempenho do cérebro (cujo funcionamento é levado a cabo por redes neuronais biológicas) com o desempenho de um computador sequencial para chegarmos à conclusão de que o cérebro é muito mais rápido e eficiente no processsamento de imagens. E isto apesar de o computador ser muito mais rápido do que um neurónio tomado individualmente. A grande vantagem do cérebro consiste no alucinante número de neurónios nele existentes e no número de conexões que entre eles se estabelecem. O facto de todos eles trabalharem em conjunto para a mesma finalidade permite-lhe ser mais rápido e fiável que um algoritmo sequêncial. Assim, as redes neuronais poderão ser mais vantajosas para a resolução deste tipo de problemas devido, principalmente, a duas características importantes: o processamento de informação em paralelo e a capacidade de generalizarem para casos não encontrados anteriormente.

O modelo sugerido por este trabalho para a obtenção de informação acerca da tridimensionalidade de uma cena poderá ser utilizado em aplicações de Realidade Virtual, nomeadamente, no domínio da Telepresença. Um dispositivo remoto de captação de imagens enviaria os dados captados para o sistema de processamento de imagens, o qual inferiria acerca da estrutura tridimensional das cenas captadas. Uma vez obtida a estrutura tridimensional, os utilizadores poderiam interagir, através de equipamento apropriado, sob duas formas básicas. Em primeiro lugar, poderiam interagir directamente no ambiente real através do comando à distância de certos equipamentos, por exemplo, *robots*. A outra forma seria uma interacção e/ou imersão em termos virtuais com a representação tridimensional obtida.

As possibilidades de aplicação de um sistema deste tipo são imensas. Ambientes que de outra forma seriam inacessíveis (por serem distantes ou perigosos) poderão estar ao alcance dos operadores deste sistema para com eles interagirem. Por exemplo, poderão ser recriados ambientes para vários tipos de testes e ensaios, manipulação de objectos em ambientes radioactivos, recriação de locais arqueológicos que não possam ser alterados, realização de reuniões virtuais, etc. Como facilmente se conclui, o sistema

proposto constitui um Sistema de Informação com grandes potencialidades e de grande utilidade.

4.2. Redes Neuronais

Neste sub-capítulo será efectuada uma breve análise da teoria das redes neuronais. A expressão "redes neuronais" é utilizada ao longo deste trabalho como forma abreviada da expressão completa "redes neuronais artificiais", em oposição às suas homólogas biológicas. Em primeiro lugar, será levada a cabo uma breve síntese dos principais desenvolvimentos e contributos verificados nesta área de investigação. De seguida, será explicado, de forma muito simplificada, o funcionamento dos neurónios biológicos e dos neurónios artificiais e serão referidas algumas arquitecturas de redes neuronais mais usuais. Por fim, serão feitas algumas considerações acerca das redes neuronais e serão dados alguns exemplos das aplicações das redes neuronais.

4.2.1. Síntese Histórica

A moderna teoria das redes neuronais deve a sua existência às tentativas desenvolvidas há várias décadas atrás pelos investigadores para conseguirem entender o funcionamento do cérebro e, em particular, das células que o compõem, os neurónios. Foram feitas tentativas para modelizar e representar esquematicamente o funcionamento dos neurónios. Uma das primeiras, e que constituíu a base da neurocomputação, foi desenvolvida por **McCulloch e Pitts (1943)**. Estes autores tiveram o mérito de serem os primeiros a tratarem o cérebro como um organismo computacional.

Posteriormente, **Rosenblatt (1958)** apresenta uma estrutura neuronal a que deu o nome de ***perceptron***. O seu objectivo era simular a computação neuronal por forma a executar tarefas complexas e ilustrar algumas das propriedades fundamentais dos sistemas inteligentes em geral. Rosenblatt acreditava que as conexões que se desenvolvem nas redes neuronais biológicas são devidas em grande parte a elementos aleatórios. Assim, a ferramenta de análise mais apropriada seria a teoria das probabilidades. Esta sua crença levou-o a desenvolver a teoria da **separabilidade**

estatística, que utilizou para caracterizar as propriedades destas redes aleatoriamente interconectadas.

O *perceptron* era um dispositivo com capacidade para aprender, o que lhe permitia classificar e fazer uma diferenciação entre padrões. Na sua configuração inicial era incapaz de distinguir padrões com interesse. No entanto, através de um processo de treino, poderia aprender esta capacidade. Esse treino consistia em introduzir vários *inputs* na rede e verificar se esta activava a resposta correcta. Se a resposta fosse correcta, era aumentada a importância (peso) computacional dos neurónios que contribuiam para a resposta obtida. Se a resposta fosse errada, o seu peso diminuiria. O processo de treino continuaria até se obter consistentemente respostas correctas. A precisão do *perceptron* diminuia com o aumento do número de padrões que tinha que aprender.

O trabalho de Rosenblatt permitiu-lhe provar um resultado importante conhecido como o **teorema da convergência do *perceptron***. Este teorema estabelece que se uma classificação poder ser aprendida pelo *perceptron*, então o processo de treino atrás descrito garante que ela de facto será aprendida num número finito de ciclos de treino.

Por volta da mesma altura, redes semelhantes ao *perceptron* foram inventadas por **Widrow e Hoff (1960)**, que lhes deram o nome de ***adalines***.

Na altura em que foi apresentado, o *perceptron* causou grande controvérsia, em parte devido às grandes expectativas que foram colocadas quanto à sua utilidade e desempenho. O seu bom desempenho estava sujeito a algumas condições que tinham que ser respeitadas. Algum tempo depois, um trabalho publicado por **Minsky e Papert (1969)** apresenta uma análise detalhada das capacidades e limitações do *perceptron*. Estes autores levantam a questão, correcta, de que o *perceptron* era indicado apenas para certos tipos de problemas, colocando uma grande restrição na sua aplicabilidade. Muitos consideram que este trabalho foi um rude golpe neste campo de investigação.

Como resultado destes factos, o campo das redes neuronais foi praticamente abandonado, excepto por alguns investigadores mais acérrimos. No entanto, no ínicio da década de 80, o interesse nesta área voltou a reacender-se de forma muito acentuada devido a uma série de factores. Em primeiro lugar, os avanços tecnológicos verificados nos computadores pessoais e *main-frames* permitiram aos investigadores simularem as suas redes neuronais e experimentarem novas idéias. Em segundo lugar, verificou-se que, apesar da velocidade de processamento dos computadores ser muito superior à do

cérebro, para certos tipos de problemas este consegue dar uma resposta muito mais rápida e eficaz. Entre esses problemas temos, por exemplo, o reconhecimento de objectos. Os computadores apenas são mais rápidos em tarefas que exijam simplesmente operações aritméticas. Assim, procurou-se desenvolver métodos de processamento em paralelo, tentando incorporar características de funcionamento do cérebro.

Por outro lado, apareceram novos trabalhos que despertaram novo interesse nas redes neuronais. Destaca-se, em primeiro lugar, a investigação desenvolvida por **Hopfield (1982)** no domínio das redes neuronais com conexões simétricas. Anteriormente, estas redes tinham sido postas de parte pelo facto de não serem semelhantes às redes neuronais biológicas. Hopfield teve o mérito de se distanciar das restrições biológicas, tendo descoberto um conjunto de propriedades e usos para as redes neuronais simétricas. Para além disso, também introduziu o conceito de **função de energia**. O trabalho de Hopfield provocou uma explosão neste domínio de investigação, conduzindo a uma série de avanços que permitiram utilizar estas redes como instrumentos para a resolução de problemas de optimização, tendo sido, por exemplo, aplicadas na resolução de instâncias de dimensão reduzida do Problema do Caixeiro Viajante.

Um dos desenvolvimentos mais importantes da década de 80 foi a introdução de um algoritmo, denominado ***back-propagation***, destinado a ajustar os pesos que conectam as várias camadas de unidades de processamento numa rede do tipo *perceptron*. Nesta área, é de destacar o trabalho de investigação de **Rumelhart, Hinton e Williams (1986)**. Apesar de não ser um algoritmo generalista, capaz de ensinar uma tarefa computacional arbitrária a uma rede neuronal, é capaz de resolver muitos problemas que os *perceptrons* na sua versão simplificada não eram.

Actualmente, muito do trabalho de investigação é desenvolvido com a finalidade de conseguir melhoramentos em relação a estes avanços verificados na década passada.

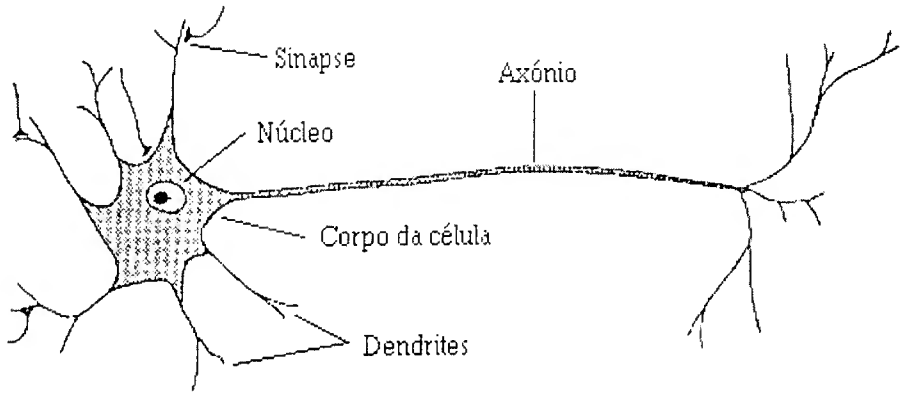
Esta breve resenha histórica foi desenvolvida tendo por base um trabalho de **Hertz, Krogh e Palmer (1991)**.

4.2.2. Neurónios Biológicos

Apesar de as redes neuronais artificiais não retratarem de forma realista o que se passa no cérebro, desde sempre que os investigadores se socorreram de conhecimentos da neurociência. Por isso, convém dar uma pequena explicação acerca do funcionamento dos neurónios biológicos. Esta explicação será feita em termos muito simplificados, pois, na realidade, os processo biológicos são muito mais complexos.

O cérebro humano é composto por aproximadamente 10^{11} neurónios de vários tipos. A Figura 4.1 ilustra um desses tipos de neurónio.

Figura 4.1 - Representação esquemática de um neurónio



Nota: Extraído de **Hertz, Krogh e Palmer (1991)**

Redes de fibra nervosa chamadas dendrites encontram-se ligadas ao corpo da célula, onde está situado o seu núcleo. Estendendo-se a partir do corpo da célula encontra-se uma única fibra longa que tem o nome de axónio e que se ramifica em várias terminações. O fim destas terminações estabelecem ligações com as dendrites e com os corpos de outros neurónios. Essas ligações têm o nome de junções sinápticas ou, simplesmente, sinapses. É nas sinapses que se dá a transmissão de sinais de neurónio para neurónio. O axónio de um neurónio típico estabelece alguns milhares de sinapses com outros neurónios. Por sua vez, cada neurónio é capaz de receber cerca de 10 000 ligações sinápticas a partir de outros neurónios.

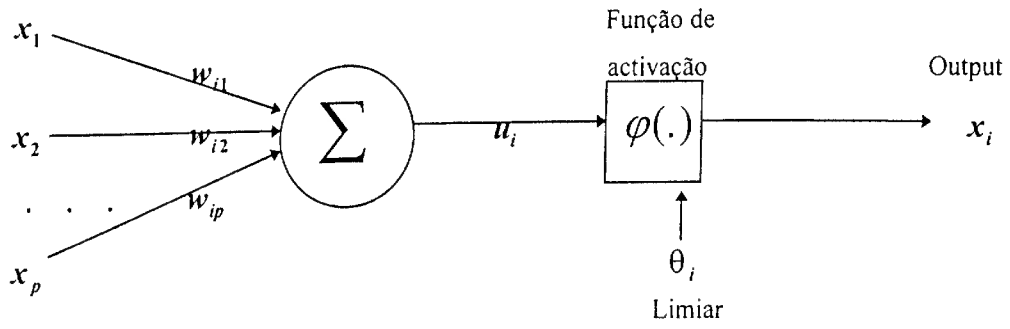
A transmissão de sinais de um neurónio para outro, através das sinapses, é um processo químico complexo. Neste processo, substâncias denominadas

neurotransmissores são libertadas do lado emissor da sinapse, as quais vão alterar a permeabilidade da parte receptora da sinapse a certas espécies de iões. Um influxo de iões positivos farão despolarizar o potencial de repouso da célula, provocando um efeito excitador. Se o influxo for de iões negativos, dá-se um efeito de hiperpolarização, sendo este efeito inibidor. Ambos estes efeitos são efeitos locais, que se propagam para o corpo da célula, onde vão sendo acumulados. Se a soma for superior a um certo limite, gera-se um potencial de acção, o qual será enviado ao longo do axónio. Diz-se então que a célula foi activada. O impulso será dividido pelas várias terminações do axónio, até às sinapses com outra células. Depois de activada, a célula tem que esperar durante um certo tempo, o denominado período refractário, antes de poder ser novamente activada. O período refractário limita a frequência de transmissão de impulsos a cerca de 1 000 por segundo.

4.2.3. Neurónios Artificiais

Um neurónio artificial é uma unidade processadora de informação fundamental para o funcionamento de uma rede neuronal. **Haykin (1994)** apresenta uma representação esquemática de um neurónio artificial, a qual se encontra adaptada na Figura 4.2.

Figura 4.2 - Modelo de um neurónio artificial



Nota: Adaptado de **Haykin (1994)**

Neste modelo é possível identificar três elementos básicos do neurónio:

1 - Um conjunto de sinapses que transmitem sinais ao neurónio, sendo cada uma dessas sinapses caracterizada por um peso. Especificamente, um sinal de *input* x_j na sinapse j, conectada ao neurónio i, é multiplicado pelo peso sináptico w_{ij}. O índice i refere-se ao neurónio em questão, enquanto j identifica qual o neurónio que lhe está a fornecer o *input* e, consequentemente, qual o peso que está associado a essa conexão. No caso da figura temos $j = 1, 2, ..., p$ neurónios conectados ao i - ésimo neurónio. O peso w_{ij} representa a "intensidade" da sinapse. Se o peso for positivo então essa sinapse é excitadora. Se for negativo, a sinapse em questão é inibidora. Caso o peso seja nulo, então não existe conexão entre os dois neurónios em questão.

2 - Um somatório, que soma os sinais de *input*, ponderados pelos pesos associados às respectivas sinapses.

3 - Uma função de activação que tem por finalidade limitar a amplitude de valores do *output* do neurónio. Normalmente, a amplitude normalizada do *output* está limitada ao intervalo $[0,1]$ ou, alternativamente, $[-1,1]$. Nalguns casos (neurónios binários), limita-se a assumir os valores 0 (inactivo) e 1 (activo).

O modelo do neurónio da Figura 4.2 contém ainda um *limiar* θ_k aplicado externamente. Esse limiar tem por efeito diminuir o *input* líquido da função de activação. Por outro lado, o *input* líquido da função de activação poderá ser aumentado se for utilizado o valor simétrico do limiar.

O i - ésimo neurónio poderá ser descrito em termos matemáticos da seguinte forma:

$$u_i = \sum_{j=1}^{p} w_{ij} x_j \tag{4.1}$$

e por

$$x_i = \varphi(u_i - \theta_i) \tag{4.2}$$

onde $x_1, x_2, ..., x_p$ são os sinais de *input*; $w_{i1}, w_{i2}, ..., w_{ip}$ são os pesos sinápticos do neurónio i; u_i é o *output* do somatório; θ_i é o limiar; $\varphi(.)$ é a função de activação e x_i

é o sinal de *output* do neurónio. A utilização do limiar θ_i tem o efeito de aplicar uma transformação de afinamento ao *output* u_i:

$$v_i = u_i - \theta_i \tag{4.3}$$

onde v_i é o potencial de activação do neurónio, ao qual irá ser de seguida aplicada a função de activação $\varphi(.)$ para se obter o *output* do neurónio.

4.2.4. Arquitecturas de Redes Neuronais

Um conjunto de neurónios artificiais interconectados formam uma rede neuronal artificial. A potência computacional da rede irá depender do número de neurónios e do número de conexões estabelecidas. Quanto maior for o seu número, maior será a sua potência computacional. Naturalmente, a estrutura da rede terá que ser estabelecida de acordo com o problema para o qual se pretende encontrar uma solução.

Quanto à forma como os neurónios se encontram dispostos e à forma como se estabelecem as conexões entre neurónios, temos dois tipos de arquitectura: arquitectura ***feedforward*** e ***feedback***, as quais se encontram ilustradas na Figura 4.3 e na Figura 4.4.

Figura 4.3 - Arquitectura *feedforward*

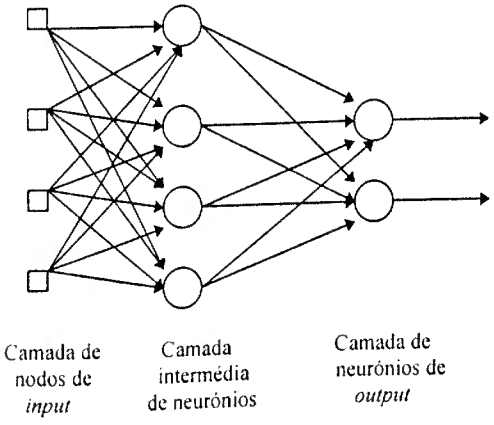
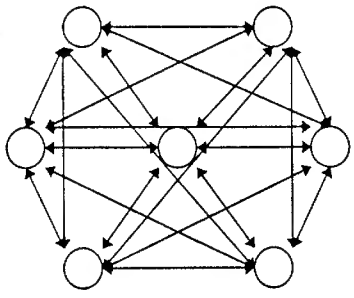


Figura 4.4 - Arquitectura *feedback*



No caso da arquitectura *feedforward*, a rede é formada por várias camadas de neurónios. A primeira camada é constituída pelos nodos de *input* (não são considerados neurónios porque nesta camada não é efectuada nenhuma computação) que vão fornecer

os dados de activação aos neurónios da camada intermédia (nodos de computação). Os *outputs* desta segunda camada serão utilizados como *input* da camada seguinte (que, no caso desta figura, constitui a última camada). Existe, portanto, um processo sequencial. Cada camada de neurónios recebe como *input* somente sinais da camada anterior. Para além disso, as conexões são unidireccionais, ou seja, o fluxo de sinais dirige-se numa só direcção. O conjunto de sinais de *output* da última camada de neurónios é a resposta encontrada pela rede ao conjunto de dados introduzido na primeira camada.

A rede neuronal da Figura 4.3 diz-se totalmente conectada uma vez que todos os nodos em cada camada se encontram conectados a todos os nodos da camada seguinte. Caso contrário, se não existirem algumas dessas conexões, diz-se que a rede é parcialmente conectada. Este último caso verifica-se normalmente quando existe informação *a priori* sobre os dados de activação da rede.

O caso da arquitectura *feedback* distingue-se da anterior pelo facto de se estabelecer pelo menos um circuito fechado entre nodos, ou seja, os sinais já não seguem uma só direcção. Existirão sinais a seguir para a frente e outros para trás. Poderá verificar-se o caso em que o *output* de um neurónio sirva de *input* para ele próprio e/ou para os outros neurónios dessa camada ou de camadas anteriores. A Figura 4.4 ilustra o caso extremo em que todos os neurónios se encontram conectados a todos os outros.

No caso deste tipo de arquitectura, o processamento dos sinais é efectuado de forma não sequencial e assíncrona, ao contrário da arquitectura *feedforward*. A presença de circuitos de *feedback* afectam profundamente a capacidade de aprendizagem da rede neuronal e o seu desempenho. Para além disso, o facto de existirem circuitos confere à rede uma dinâmica não linear.

Uma das grandes vantagens das redes neuronais consiste no facto de poderem ser ensinadas a resolver determinado problema através de um processo de treino. Esse treino consiste em introduzir na rede os dados referentes ao problema em questão e, depois de obtida a resposta, fornecer-lhe *feedback* sobre a correcção ou incorrecção dessa resposta. Mediante esse *feedback*, a rede ir-se-á adaptar. Esse processo de adaptação é conseguido através da lei de aprendizagem, a qual consiste numa equação que modifica todos ou alguns dos pesos associados a cada neurónio artificial. É esta equação que permite à rede adaptar-se aos exemplos fornecidos como treino e efectuar as modificações necessárias nos pesos por forma a conseguir chegar às respostas pretendidas. Eventualmente,

chegará um ponto em que a rede atingirá uma estabilidade que lhe permita identificar as respostas correctas, podendo então ser utilizada para resolver casos novos desse tipo de problema com os quais não se tenha deparado no processo de treino. Portanto, é através da modificação dos pesos que a rede neuronal é capaz de aprender e adaptar-se. Existem alguns tipos de redes neuronais em que não é necessário qualquer tipo de treino prévio. A rede modificará automaticamente os seus pesos até que chegue a um ponto de estabilidade. Algumas redes até nem têm capacidade para aprender, como é o caso das redes de Hopfield, as quais são utilizados para resolução de problemas de optimização.

Outro elemento fundamental na rede neuronal é a função de transferência. Esta consiste numa fórmula matemática que, entre outras coisas, determina o sinal de *output* dos neurónios em função dos seus sinais de *input* e dos pesos. Para além disso, inclui também a lei de aprendizagem dos neurónios.

Por último, temos a função de escalonamento, que determina se e com que frequência um determinado neurónio deverá aplicar a sua função de transferência.

4.2.5. Algumas Considerações Sobre Redes Neuronais

Em resumo, uma rede neuronal, ao contrário dos processos sequênciais de resolução de problemas, é um modelo que gera ele próprio as suas regras internas para a interpretação dos dados que são introduzidos. Essas regras serão aperfeiçoadas através da comparação dos seus resultados com os exemplos. Por meio de um processo de tentativas, a rede aprende a desempenhar determinada tarefa. Isto torna difícil (ou mesmo impossível) conhecer as regras utilizadas pela rede para resolver os problemas que lhe são colocados.

Cada neurónio da rede é completamente auto-suficiente, funcionando independentemente do processamento que decorre no interior dos seus vizinhos. Isto torna a rede neuronal muito potente e resistente a falhas ou erros eventuais de neurónios. O facto de um neurónio deixar de "funcionar", geralmente, não impedirá a rede de chegar a uma resposta correcta. O mesmo se passa com o cérebro, uma vez que, apesar de ao longo do tempo ir perdendo neurónios e conexões, o seu bom desempenho não é significativamente afectado a curto prazo.

Por outro lado, todos os neurónios afectam conjuntamente o desempenho da rede, uma vez que o *output* de cada neurónio se torna no *input* de muitos outros. A topologia das conexões entre os neurónios influencia o tipo de tarefas de processamento de informação que a rede pode desempenhar, uma vez que determina a informação que cada neurónio recebe dos outros. A importância de cada conexão no processamento é determinada pelo peso que lhe está associado: quanto maior for o peso maior será o contributo dessa conexão para o *output* do neurónio em questão.

Se os conhecimentos de uma rede neuronal se encontram incorporados nos seus pesos, uma das questões fundamentais que se coloca é a determinação desses pesos. De uma forma geral, a rede pode ser ensinada a realizar determinada tarefa através de ajustes iteractivos dos pesos. Normalmente, uma rede neuronal aprende a processar informação através de um processo de treino supervisionado ou de treino não supervisionado. No primeiro caso, são fornecidos à rede os dados de *input* e o *output* desejado (a resposta correcta a que ela deveria chegar). Depois de cada tentativa, a rede compara o seu *output* com a resposta correcta, procura minimizar as diferenças através da modificação dos pesos e tenta novamente até que o erro no *output* atinja um valor aceitável. Neste tipo de treino, temos aínda o caso particular em que não é fornecida a resposta correcta mas somente lhe é indicado se a resposta a que chegou está correcta ou não. Em ambos os casos, o treino é levado a cabo até que a rede atinja um estado estável, ou seja, quando já não existam modificações significativas nos pesos.

No caso do treino não supervisionado, o objectivo da aprendizagem não é definido em termos de identificação de respostas correctas. A única informação disponível diz respeito a correlações entre os dados de *input*, devendo a rede criar categorias a partir destas correlações e produzir *outputs* correspondentes à categoria de *input*.

Depois de terminado o processo de treino (quando exista estabilidade na rede), a lei de aprendizagem pode ser "desligada". Desta forma, os pesos não serão actualizados com a introdução de novos dados, o que poderá ser útil para aumentar a velocidade de processamento da rede. Esta decisão deverá ser tomada de acordo com a aplicação a que a rede se destina.

O facto de as redes neuronais poderem aprender através de um processo de treino é muito atractivo, pois, em vez de especificar todos os detalhes e regras referentes a certa

computação, bastará compilar um conjunto de exemplos e treinar uma rede neuronal na execução dessa computação. A capacidade que as redes neuronais têm para aprender permite-lhe generalizar, ou seja, produzir *outputs* aceitáveis face a *inputs* nunca antes encontrados. Isto é particularmente útil nos casos de problemas em que é difícil conhecer-se antecipadamente as regras necessárias para a sua resolução, como por exemplo no caso de sistemas periciais.

A investigação na área das redes neuronais conduziu a significativos avanços no campo da Inteligência Artificial. Tais avanços permitiram desenvolver uma série de aplicações práticas, entre as quais temos, por exemplo, reconhecimento de caracteres escritos à mão (reconhecimento de assinaturas e códigos postais escritos à mão), reconhecimento de estruturas de protéinas, reconhecimento de voz e imagem, filtragem de ecos nas redes de telecomunicações, radar, sonar, sismologia e compressão de sons e imagens.

Sharda (1994) apresenta uma série de aplicações das redes neuronais nos campos da Gestão/Finanças, Investigação Operacional e Estatística. Nomeadamente, neste último caso, as redes neuronais têm sido aplicadas na análise de regressão, previsão de séries cronológicas e classificação. No campo da Gestão/Finanças, foram desenvolvidos algoritmos para previsão de falências de empresas, previsão de falências de bancos, previsão de preços de acções, cotação de obrigações, alocação de activos, prevenção de fraudes, previsão de fusões de empresas, previsão de empresas-alvo de processos de aquisição e classificação do risco de pedidos de concessão de crédito. Quanto ao campo da Investigação Operacional, o interesse pelas redes neuronais como instrumento para a resolução de problemas de optimização desperta com um artigo publicado por **Hopfield e Tank (1985)**. Desde então, os investigadores têm aplicado as redes neuronais à maioria das áreas da investigação operacional, das quais se destaca a optimização combinatorial, programação linear e não linear, selecção de rotas de veículos e problemas de afectação.

As redes neuronais podem ser encaradas como um Sistema de Informação, uma vez que apresentam características e mecanismos que lhes permitem processar informação, armazená-la e recarregá-la. A associação da área das redes neuronais com a tecnologia VLSI permitirá ultrapassar as limitações inerentes aos actuais computadores,

cujo processamento é sequêncial. Já há alguns anos que se começou a desenvolver e comercializar *hardware* que implementa o paralelismo das redes neuronais, os chamados neurocomputadores. Esta união permitirá desenvolver sistemas que, apesar de tirarem partido do paralelismo das redes neuronais, serão suficientemente rápidos para o desenvolvimento de aplicações em tempo real. Portanto, quando os avanços a nível de *hardware* permitirem implementar redes neuronais massivas, testemunharemos, certamente, uma nova revolução nas Tecnologias de Informação e nos Sistemas de Informação.

4.3. Formalização do Problema Utilizando Redes Neuronais

Neste sub-capítulo vão ser apresentados os modelos baseados em redes neuronais necessários para a obtenção da estrutura tridimensional dos objectos a partir da análise de uma sequência de imagens bidimensionais. Primeiro, é apresentado um modelo de rede de Hopfield para o estabelecimento de correspondências entre pontos característicos de duas imagens. Depois, é apresentada uma rede neuronal para a inferência sobre a estrutura tridimensional dos objectos a partir das correspondências estabelecidas.

4.3.1. Extracção de Características e Estabelecimento de Correspondências

Para o estabelecimento de correspondências entre pontos característicos irá ser adoptado um modelo de rede neuronal de Hopfield desenvolvido por **Nasrabadi e Choo (1992)**. Antes que o modelo possa ser aplicado, será necessário extrair os pontos característicos das imagens em relação às quais se pretender estabelecer correspondências. Para essa finalidade, Nasrabadi e Choo propõem a utilização da técnica desenvolvida por **Moravec (1977)** e **(1981)**. Essa técnica detecta pontos característicos (cantos de objectos) através da análise das variações na intensidade da luminosidade das imagens. A esses pontos Moravec dá o nome de pontos de interesse. A técnica desenvolvida por este investigador encontra-se suncintamente explicada no sub-capítulo 3.2. Nasrabadi e Choo decidiram utilizar esta técnica pelo facto de ser de

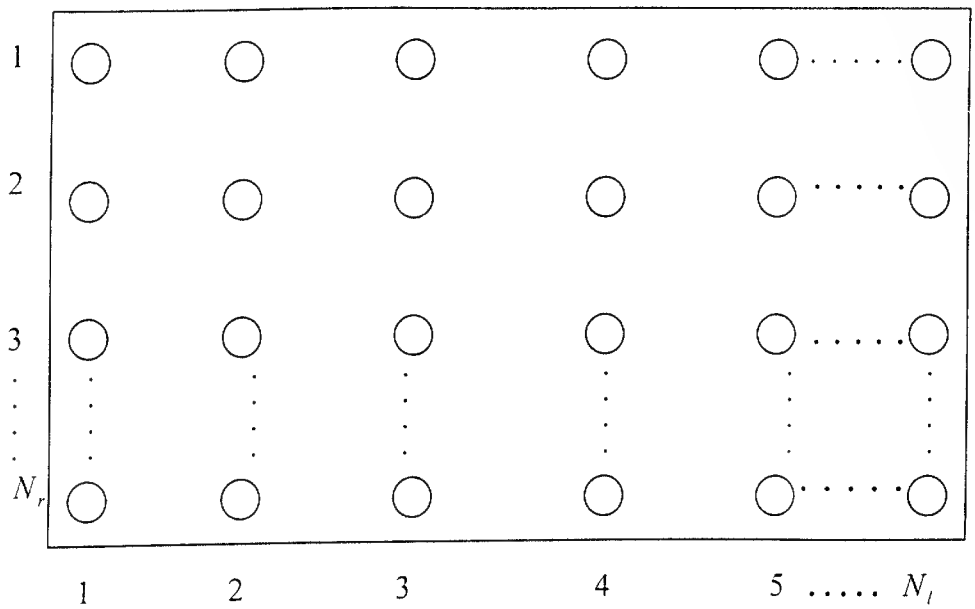
implementação computacional eficiente, quando comparado com outros detectores de cantos mais complicados.

A formalização do problema do estabelecimento de correspondências utilizando uma rede neuronal poderá ser levada a cabo através da minimização de uma função de custo. Essa função de custo irá incluir todas as restrições que deverão ser colocadas para possibilitar a obtenção da solução óptima. Essa minimização da função de custo será conseguida através da aplicação de uma rede de Hopfield, a qual é apropriada para problemas de optimização.

A função de custo atrás referida traduz-se na função de Lyapunov (também denominada função de energia), a qual representa o comportamento colectivo da rede. Quanto aos pesos associados às conexões entre os neurónios, reflectem as restrições impostas ao problema do estabelecimento de correspondências. A rede irá tomar uma decisão tendo por base os contributos de todos os neurónios que a compõem. Cada neurónio recebe e fornece sinais a todos os outros neurónios (não existe *feedback* de neurónios para si próprios). Esta troca de informação permitirá à rede convergir para um estado estável, o qual será conseguido quando a função de energia se encontrar no seu mínimo, permitindo-lhe assim tomar uma decisão.

Para o estabelecimento de correspondências, será utilizado um modelo bidimensional de rede de Hopfield. A rede é constituída por $N_l \times N_r$ neurónios, onde N_l e N_r representam o número total de pontos de interesse nas imagens esquerda e direita, respectivamente. Na Figura 4.5 encontra-se representada a estrutura de neurónios da rede neuronal que irá ser utilizada.

Figura 4.5 - Estrutura da rede neuronal de Hopfield

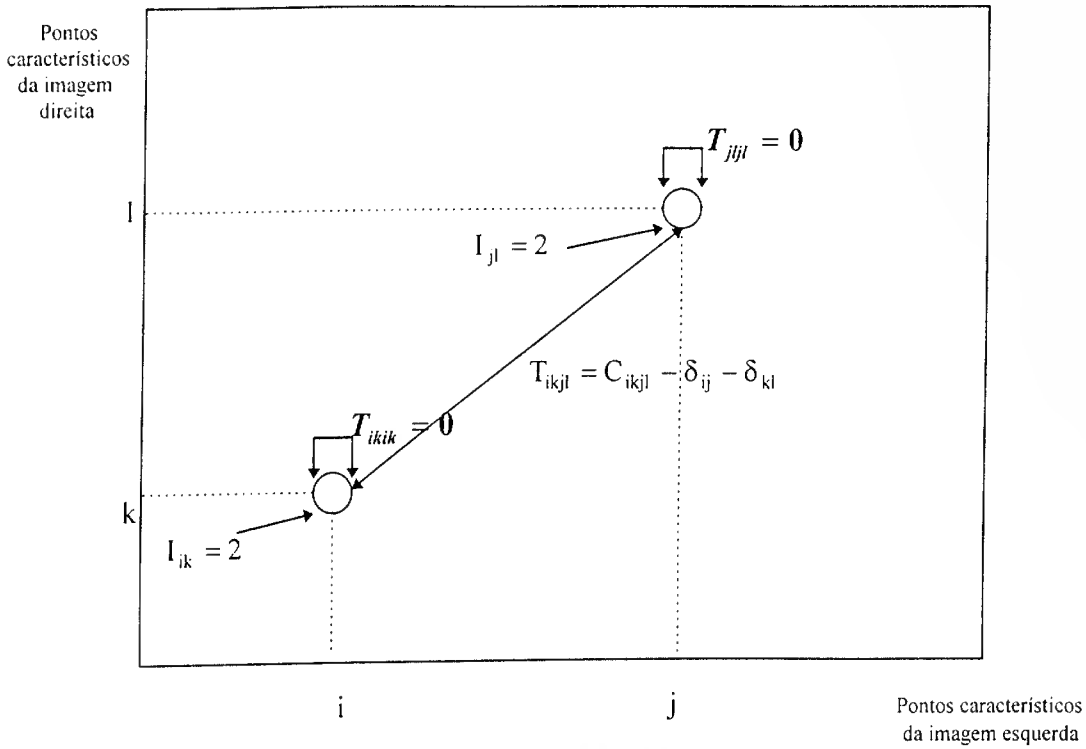


Cada neurónio pode assumir dois estados: ligado ou desligado. Se determinado neurónio se encontrar ligado, isso indica uma possível correspondência entre o ponto de interesse da imagem esquerda e o ponto de interesse da imagem direita que se encontram associados a esse neurónio. Em cada linha e em cada coluna apenas deverá existir um neurónio ligado, como resultado da restrição de unicidade. Esta restrição obriga a que cada ponto de interesse na imagem esquerda tenha apenas um ponto de interesse correspondente na imagem direita e vice-versa.

A restrição da unicidade é uma boa forma de combater os efeitos nocívos da oclusão e das sombras. Estes dois fenómenos poderão fazer com que pontos que estejam presentes numa imagem não estejam visíveis na outra, existindo, portanto, pontos que ficarão sem correspondências. Estes pontos poderão ser correspondidos de forma errada a outros e conduzir ao estabelecimento de correspondências múltiplas. A restrição da unicidade permitirá contrariar este problema.

Quanto às conexões entre neurónios, cada neurónio recebe e envia sinais para todos os outros, excepto para si próprio. A Figura 4.6 ilustra as conexões que se estabelecem entre quaisquer dois neurónios (entre os neurónios *ik* e *jl*, no caso da figura) e os pesos que lhes estão associados.

Figura 4.6 - Conexões entre os neurónios



Nota: Adaptado de **Nasrabadi e Choo (1992)**

O peso associado à conexão entre os neurónios *ik* e *jl* é representado por T_{ikjl}, que é igual a T_{jlik} uma vez que as conexões entre neurónios são bidireccionais. Para fazer com que não exista conexão de cada neurónio para ele próprio é necessário dar o valor zero ao peso T_{ikik}. Quanto a I_{ik}, representa o *input* inicial para cada um dos neurónios (mais à frente será demonstrado porque deve ser igual a dois).

Para uma rede de Hopfield bidimensional, a função de Lyapunov (ou função de energia) é dada pela seguinte expressão:

$$E = -\frac{1}{2}\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} T_{ikjl}V_{ik}V_{jl} - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r} I_{ik}V_{ik} \qquad (4.4)$$

onde V_{ik} e V_{jl} representam os estados binários dos neurónios *ik* e *jl*, respectivamente, os quais podem assumir os valores 1 (activo) ou 0 (inactivo).

Uma modificação de ΔV_{ik} no estado do neurónio *ik* irá provocar uma modificação de ΔE_{ik} na energia:

$$\Delta E_{ik} = -\left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} T_{ikjl}V_{jl} + I_{ik}\right]\Delta V_{ik} \qquad (4.5)$$

A equação (4.5) traduz a dinâmica da rede neuronal. Hopfield provou que essa dinâmica é sempre negativa, tendo também determinado uma regra estocástica de actualização dos neurónios:

$$V_{ik} \longrightarrow 0 \quad \text{se} \quad \left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} T_{ikjl}V_{jl} + I_{ik}\right] < 0$$

$$V_{ik} \longrightarrow 1 \quad \text{se} \quad \left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} T_{ikjl}V_{jl} + I_{ik}\right] > 0$$

$$\text{sem alteração se} \quad \left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} T_{ikjl}V_{jl} + I_{ik}\right] = 0 \qquad (4.6)$$

Para resolver o problema do estabelecimento de correspondências será necessário incorporar na função de energia as restrições a aplicar para que seja possível encontrar uma solução. A função que será minimizada pela rede é a seguinte:

$$E = -\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} C_{ikjl}P_{ik}P_{jl} + \sum_{i=1}^{N_l}\left(1-\sum_{k=1}^{N_r}P_{ik}\right)^2 + \sum_{k=1}^{N_r}\left(1-\sum_{i=1}^{N_l}P_{ik}\right)^2 \qquad (4.7)$$

O primeiro termo nesta equação representa o grau de compatibilidade de uma correspondência entre um par de pontos (i,j) da imagem esquerda e um par de pontos (k,l) na imagem direita. Os segundo e terceiro termos da equação são os responsáveis pela imposição da restrição da unicidade, devendo a soma das probabilidades (que representam o estado dos neurónios) em cada linha e em cada coluna dar um total de 1 (para que exista uma única correspondência em cada linha e em cada coluna). Cada uma das probabilidades representa uma medida de correspondência entre um ponto característico da imagem esquerda e um ponto característico da imagem direita.

De seguida, a equação (4.7) será desenvolvida por forma a se assemelhar mais com a equação de energia (4.4). Comecemos por simplificar os dois últimos termos:

$$\sum_{i=1}^{N_l}\left(1-\sum_{k=1}^{N_r}P_{ik}\right)^2=\sum_{i=1}^{N_l}(1)+\sum_{i=1}^{N_l}\left(\sum_{k=1}^{N_r}P_{ik}\right)^2-2\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}P_{ik}\Leftrightarrow \tag{4.8}$$

$$\Leftrightarrow\sum_{i=1}^{N_l}\left(1-\sum_{k=1}^{N_r}P_{ik}\right)^2=\sum_{i=1}^{N_l}(1)+\sum_{i=1}^{N_l}\left(\sum_{k=1}^{N_r}\sum_{l=1}^{n_r}P_{ik}P_{il}\right)^2-\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}2P_{ik}\Leftrightarrow \tag{4.9}$$

$$\Leftrightarrow\sum_{i=1}^{N_l}\left(1-\sum_{k=1}^{N_r}P_{ik}\right)^2=N_l+\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{l=1}^{N_r}P_{ik}P_{il}-\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}2P_{ik} \tag{4.10}$$

Substituíndo, em (4.10), $P_{ik}P_{il}$ por $P_{ik}P_{jl}\delta_{ij}$, teremos:

$$\Leftrightarrow\sum_{i=1}^{N_l}\left(1-\sum_{k=1}^{N_r}P_{ik}\right)^2=N_l+\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}P_{ik}P_{jl}\delta_{ij}-\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}2P_{ik} \tag{4.11}$$

De forma semelhante, também se pode simplificar o terceiro termo:

$$\sum_{k=1}^{N_r}\left(1-\sum_{i=1}^{N_l}P_{ik}\right)^2=\sum_{k=1}^{N_r}(1)+\sum_{k=1}^{N_r}\left(\sum_{i=1}^{N_l}P_{ik}\right)^2-2\sum_{k=1}^{N_r}\sum_{i=1}^{N_l}P_{ik}\Leftrightarrow \tag{4.12}$$

$$\Leftrightarrow\sum_{k=1}^{N_r}\left(1-\sum_{i=1}^{N_l}P_{ik}\right)^2=\sum_{k=1}^{N_r}(1)+\sum_{k=1}^{N_r}\left(\sum_{i=1}^{N_l}\sum_{j=1}^{N_l}P_{ik}P_{ij}\right)^2-\sum_{k=1}^{N_r}\sum_{i=1}^{N_l}2P_{ik}\Leftrightarrow \tag{4.13}$$

$$\Leftrightarrow\sum_{k=1}^{N_r}\left(1-\sum_{i=1}^{N_l}P_{ik}\right)^2=N_r+\sum_{k=1}^{N_r}\sum_{i=1}^{N_l}\sum_{j=1}^{N_l}P_{ik}P_{ij}-\sum_{k=1}^{N_r}\sum_{i=1}^{N_l}2P_{ik} \tag{4.14}$$

Substituíndo, em (4.14), $P_{ik}P_{ij}$ por $P_{ik}P_{jl}\delta_{kl}$, teremos:

$$\Leftrightarrow\sum_{k=1}^{N_r}\left(1-\sum_{i=1}^{N_l}P_{ik}\right)^2=N_r+\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}P_{ik}P_{jl}\delta_{kl}-\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}2P_{ik} \tag{4.15}$$

Assim sendo, podemos voltar a equacionar a expressão (4.7) da seguinte forma:

$$E = \left(N_l + N_r\right) + \sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}\left(-C_{ikjl}P_{ik}P_{jl} + P_{ik}P_{jl}\delta_{ij} + P_{ik}P_{jl}\delta_{kl}\right) - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r}4P_{ik} \quad (4.16)$$

$$\Leftrightarrow E - \left(N_l + N_r\right) = \sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}\left(-C_{ikjl} + \delta_{ij} + \delta_{kl}\right)P_{ik}P_{jl} - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r}4P_{ik} \Leftrightarrow \quad (4.17)$$

$$\Leftrightarrow \frac{E - \left(N_l + N_r\right)}{2} = -\frac{1}{2}\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}\left(C_{ikjl} - \delta_{ij} - \delta_{kl}\right)P_{ik}P_{jl} - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r}2P_{ik} \quad (4.18)$$

A expressão (4.18) é equivalente à função de Lyapunov de uma rede de Hopfield:

$$\frac{E - \left(N_l + N_r\right)}{2} = -\frac{1}{2}\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}T_{ikjl}V_{ik}V_{jl} - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r}I_{ik}V_{ik} \quad (4.19)$$

Assim sendo, o peso associado à conexão entre dois neurónios é dado por $T_{ikjl} = \left(C_{ikjl} - \delta_{ij} - \delta_{kl}\right)$ e o *input* inicial para cada neurónio é $I_{ik} = 2$. $V_{ik} = P_{ik}$ e $V_{jl} = P_{jl}$ representam o estado dos neurónios.

Na expressão (4.18), $\delta_{ij} = 1$ se $i = j$. Caso contrário, será igual a 0. Da mesma forma, $\delta_{kl} = 1$ se $k = l$ e, caso contrário, igual a 0. Quanto à medida de compatibilidade, será calculada pela seguinte fórmula:

$$C_{ikjlj} = \frac{2}{\left[1 + e^{\lambda(X-\theta)}\right]} - 1 \quad (4.20)$$

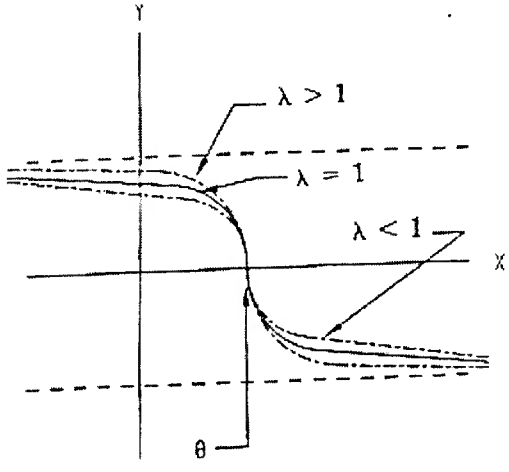
sendo X dado pela expressão:

$$X = \left[W_1|\Delta d| + W_2|\Delta D|\right]. \quad (4.21)$$

As expressões (4.20) e (4.21) mostram que a compatibilidade é analisada através de dois tipos de comparação para os dois pares de pontos correspondidos. A primeira comparação, Δd, consiste na diferença entre as disparidades dos pares de pontos correspondidos (i,k) e (j,l). Caso os pontos pertençam ao mesmo objecto, então a diferença das disparidades deverá ser pequena, assumindo que existe rígidez dos objectos e continuidade das suas superfícies. Quanto à segunda comparação, ΔD, é a diferença entre a distância de i a j e a distância de k a l, a qual é pequena quando os pontos se encontram correctamente correspondidos. W_1 e W_2 são pesos que ponderam a importância a atribuir a cada uma das comparações, devendo a sua soma ser igual a 1. Nasrabadi e Choo sugerem que se faça $W_1 = 0.4$ e $W_2 = 0.6$, uma vez que consideram ΔD mais estável do que Δd.

A expressão (4.20) é uma função não linear que faz com que a medida de compatibilidade varie entre os valores +1 e -1, encontrando-se o seu gráfico representado na Figura 4.7.

Figura 4.7 - Gráfico da função de compatibilidade



Nota: Extraído de **Nasrabadi e Choo (1992)**

O parâmetro λ estabelece a inclinação da função. Caso λ seja muito elevado, a função variará entre +1 e -1 de uma forma muito brusca. Se λ for reduzido, a função variará entre +1 e -1 de forma muito suave, podendo a compatibilidade assumir muitos valores neste intervalo. O parâmetro θ define a posição onde a função corta o eixo dos

X, devendo o seu valor ser escolhido por forma a que a medida de compatibilidade dê um valor de +1 para uma boa correspondência, (quando X, na expressão (4.21), for igual a 0). Quando X não for exactamente igual a 0 (devido a efeitos se "ruído", por exeplo), a compatibilidade deverá ser 0 e igual a -1 quando a correspondência for má. Pelas suas experiências, Nasrabadi e Choo consideram que os valores mais adequados são $\lambda = 1$ e $\theta = 10$.

Uma vez inicializada, a rede irá funcionar por forma a minimizar a sua função de energia (ou função de custo) através da acção conjunta dos seus neurónios. Uma alteração no estado de um neurónio irá provocar uma alterção na energia de ΔE_{ik}, dada por:

$$\Delta E_{ik} = -\left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}\left(C_{ikjl} - \delta_{ij} - \delta_{kl}\right)P_{jl} + 2\right]\Delta P_{ik} \qquad (4.22)$$

sendo esta expressão obtida por analogia com a expressão (4.5). A alteração na energia é sempre negativa, se os neurónios forem actualizados aleatoriamente e de forma assíncrona de acordo com a seguinte regra de actualização de Hopfield:

$$\begin{aligned}
P_{ik} &\longrightarrow 0 \quad \text{se} \quad \left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}\left(C_{ikjl} - \delta_{ij} - \delta_{kl}\right)P_{jl} + 2\right] < 0 \\
P_{ik} &\longrightarrow 1 \quad \text{se} \quad \left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}\left(C_{ikjl} - \delta_{ij} - \delta_{kl}\right)P_{jl} + 2\right] > 0 \\
\text{sem alteração} &\quad \text{se} \quad \left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}\left(C_{ikjl} - \delta_{ij} - \delta_{kl}\right)P_{jl} + 2\right] = 0
\end{aligned} \qquad (4.23)$$

A actualização dos neurónios é simulada escolhendo-se aleatoriamente um ponto de interesse i da imagem esquerda e abrindo-se uma janela na imagem direita. A janela será centrada nas coordenadas do ponto i deslocadas de um valor que corresponde a uma estimativa inicial para a disparidade. Todos os pontos característicos k dentro desta janela serão considerados como uma correspondência possível ao se lhes atribuir uma probabilidade $P_{ik} = 1$. A restrição da unicidade encarregar-se-á de fazer com que só

exista um ponto característico da imagem direita correspondente a cada ponto característico da imagem esquerda.

Na imagem esquerda também será aberta uma janela centrada em *i*, sendo todos os restantes pontos característicos dentro dela considerados como os *j* para a expressão (4.23). Para todos os pontos característicos que se situem fora dessa janela, assume-se que têm contribuição igual a zero para a compatibilidade, uma vez que se encontram distantes e provavelmente pertencem a outro objecto.

A utilização da janela permite eliminar pontos que à partida têm poucas possibilidades de serem candidatos para uma correspondência, evitando-se assim computações e perda de tempo desnecessárias.

A solução estará encontrada quando a rede estiver no mínimo da sua energia. A função de energia estará no seu mínimo quando as restrições referentes à compatibilidade e à unicidade estiverem satisfeitas. Isso acontecerá quando a rede atingir a estabilidade, ou seja, não existam alterações no estado dos neurónios. Através da análise dos estados dos neurónios saberemos que pontos da imagem direita estão correspondidos aos pontos da imagem esquerda.

A utilização de uma rede neuronal para o estabelecimento de correspondências apresenta a vantagem de essas correspondências serem estabelecidas automaticamente e de uma forma global. Isso acontece porque todos os neurónios se encontram interconectados, recebendo e fornecendo *feedback* a todos os restantes. O facto de todos os neurónios trabalharem para a mesma finalidade permite à rede ser computacionalmente potente e rápida.

4.3.2. Obtenção da Estrutura Tridimensional

As correspondências estabelecidas na fase anterior vão agora ser utilizadas no processo de inferência sobre a estrutura tridimensional dos objectos presentes na cena. Essa inferência será levada a cabo através da análise de uma sequência de imagens da cena obtidas a partir de ângulos ligeiramente diferentes.

O algoritmo que vai ser utilizado baseia-se na forma como o próprio sistema visual humano reconhece estruturas tridimensionais a partir da observação das suas projecções bidimensionais. Em primeiro lugar, necessita de utilizar sequências longas de

imagens para ter uma percepção correcta das estruturas. Quanto mais longa a sequência, maior será a certeza quanto à correcção da estrutura inferida. Em segundo lugar, apesar de serem preferidas transformações rígidas na percepção das estruturas, existe uma certa tolerância a desvios da rígidez. Por último, a percepção de estruturas parece seguir um processo iterativo de aproximação até atingir a precisão.

Portanto, a inferência sobre a estrutura tridimensional consiste num processo progressivo, realizado ao longo de um certo período, incorporando gradualmente a informação que vai chegando relativa a novas observações. Em cada momento, existe um modelo estimado da estrutura, o qual é consistente com a informação recebida até esse momento e que vai sendo actualizado sempre que seja recebida mais informação.

Para tentar solucionar o problema da inferência da estrutura tridimensional utilizando redes neuronais, vamos recorrer a um modelo proposto por **Laganière e Cohen (1995)**. O modelo proposto por estes investigadores baseia-se no Esquema da Rígidez Incremental de Ullman, tratado no sub-capítulo 3.4.

À partida, colocam-se algumas questões, as quais devem ser respondidas antes da concepção do algoritmo. Em primeiro lugar, que tipo de informação estrutural deverá o modelo interno explicitar. Dada a finalidade a que o modelo se destina (obtenção de coordenadas tridimensionais de objectos), este deverá ser representado em termos de um conjunto de distâncias Euclidianas entre pares de pontos tridimensionais característicos.

Em segundo lugar, como deverá ser modificado o modelo interno, que constitui a interpretação da estrutura até esse momento, dada uma nova imagem dessa estrutura, tirada de um ângulo diferente? Tal como no sistema visual humano, será necessário recorrer ao princípio da máxima rígidez, o qual constituirá uma restrição à forma como o modelo interno será modificado. Este princípio consiste no seguinte:

- Se uma transformação rígida do modelo interno for suficiente para explicar as projecções presentes na nova imagem, então não será feita qualquer modificação no modelo interno.
- Se nenhuma transformação rígida conseguir explicar as transformações presentes na nova imagem, então será permitida uma deformação mínima no modelo interno por forma a colocá-lo em concordância com as observações. O desvio da rígidez,

apesar de permitido, terá que ser mantido o menor possível.

Dado que o Esquema da Rígidez Incremental de Ullman desempenha uma parte fundamental no modelo de Laganière e Cohen, vamos voltar a formulá-lo matematicamente como o fizeram estes investigadores.

Representemos por $\{I(t);\ t = 1, 2, ...\}$ uma sequência de imagens captadas a partir de determinada cena. Assume-se que a cena é constituída por um número finito de pontos característicos, cujas projecções em imagens consecutivas foram previamente correspondidas. Representemos por $P_i(t) = (X_i(t), Y_i(t), Z_i(t))$ as coordenadas tridimensionais do ponto i, com $i = 1, 2, ..., N$, no momento t e por $p_i(t) = (x_i(t), y_i(t))$ a projecção do ponto i no plano da imagem.

A relação entre $P_i(t)$ e $p_i(t)$ depende de se assumir a utilização da projecção ortográfica ou da projecção central e das características da câmara. Em termos gerais, vamos representar a projecção como uma aplicação com a seguinte forma:

$$C:(X,Y,Z) \rightarrow (x,y) \tag{4.24}$$

No caso particular da projecção ortográfica, temos:

$$C(X,Y,Z) = (X,Y) \tag{4.25}$$

enquanto que no caso da projecção central, que está mais em conformidade com a realidade, teremos:

$$C(X,Y,Z) = \left(\frac{X}{\lambda}, \frac{Y}{\lambda}\right) \tag{4.26}$$

onde $\lambda = (Z + f)/f$, sendo f a distância de focagem da câmara.

$M(t-1) = \{\vec{P}_i(t-1); i = 1,..., N\}$ representa o modelo interno no momento t, sendo $\vec{P}_i(t-1)$ a posição tridimensional estimada do ponto i. Sendo $M(t-1)$ o modelo interno que reflecte a informação acumulada até ao momento t - 1, no momento t será

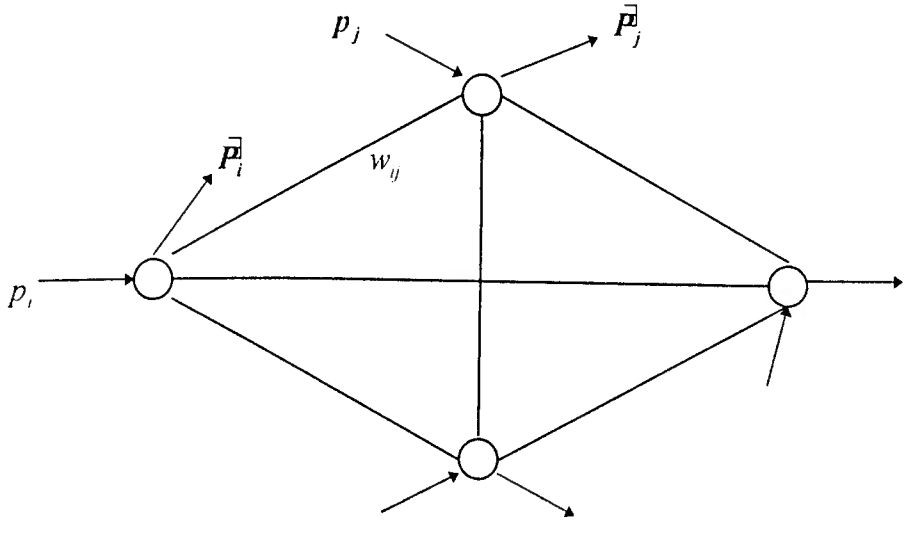
necessário actualizá-lo. Isso será conseguido aplicando-se deformações não rígidas ao modelo interno por forma a reflectir a informação contida na nova imagem $I(t)$, obtendo-se assim $M(t)$. Os desvios da rígidez seriam mantidos o menor possível através da minimização de uma função objectivo que, em termos gerais, seria dada pela seguinte expressão:

$$F = \sum_{i,j:i\neq j} \eta_{ij}\left(D\left[\vec{P}_i(t-1),\vec{P}_j(t-1)\right] - D\left[\vec{P}_i(t),\vec{P}_j(t)\right]\right)^2 \qquad (4.27)$$

em que $D\left[P_i,P_j\right]$ mede a distância entre os pontos P_i e P_j e η_{ij} são pesos atribuídos a cada par de pontos de acordo com o nível de importância (ou confiança) da sua distância mútua na função objectivo.

O princípio da máxima rígidez será de seguida implementado utilizando uma rede neuronal. Sempre que uma nova imagem $I(t)$ composta por um conjunto de pontos $p_i(t)$ seja introduzida como *input* na rede, ela irá fornecer um *output* que consiste na melhor estimativa da estrutura tridimensional, tendo em conta toda a informação que lhe foi anteriormente introduzida. A estrutura, em termos gerais, da rede neuronal que irá ser utilizada encontra-se representada na Figura 4.8.

Figura 4.8 - Estrutura da rede neuronal



Nota: Adaptado de **Laganière e Cohen (1995)**

Cada neurónio da rede representa um ponto da estrutura. O *input* do neurónio *i* é constituído pelas coordenadas $(x_i(t), y_i(t))$ do ponto $p_i(t)$ na imagem $I(t)$, enquanto que o seu *output* consistirá na estimação da posição desse mesmo ponto no espaço tridimensional. O peso $w_{ij}(t)$ associado à conexão entre os neurónios *i* e *j* corresponde à distância Euclidiana entre os pontos $\vec{P}_i(t-1)$ e $\vec{P}_j(t-1)$ na estrutura estimada no momento anterior.

A actualização do *output* de cada neurónio é feita de forma assíncrona, sendo calculado tendo por base os *outputs* dos neurónios que a ele estão conectados. Esta computação consiste em calcular a nova posição tridimensional do ponto em questão tendo por base as suas coordenadas bidimensionais na imagem e respeitando a restrição da rígidez. Por sua vez, o *output* deste neurónio vai servir de *input* para os neurónios seus vizinhos. Depois de um número suficiente de iterações, a rede atingirá a estabilidade, obtendo-se a estrutura tridimensional que mais está de acordo com todas as imagens observadas até então e que manteve as suas deformações no mínimo possível. Este processo de convergência da rede para um estado estável, mantendo os pesos $w_{ij}(t)$ constantes, tem o nome de processo de convergência intra-imagem.

Depois de atingida a estabilidade, o *output* obtido será utilizado para a elaboração de um novo modelo interno através do cálculo das distâncias entre as coordenadas tridimensionais dos vários pontos. Para além disso, estas distâncias irão também constituir o novo conjunto de pesos $w_{ij}(t+1)$ que irá ser utilizado no próximo processo de convergência intra-imagem.

A rede neuronal que vai ser utilizada apresenta as seguintes características:

- Explora a redundância presente na sequência de imagens ao calcular, em cada momento *t*, um modelo interno que resulta de toda a informação acumulada até então. Essa redundância resulta do facto de se assumir que cada ponto na imagem permance visível em mais que duas imagens. Quanto maior o número de imagens, maior será a redundância, logo a inferência acerca da estrutura será mais exacta.

- Consegue lidar com sequências com qualquer número de imagens. De facto, o

custo computacional do funcionamento da rede é independente do número de imagens utilizadas, uma vez que, em cada momento *t*, a única informação processada diz respeito ao modelo interno e à imagem actual.

- Recupera progressivamente a estrutura de uma cena e permite a existência de pequenos desvios da rígidez no modelo interno, o que está em conformidade com o próprio sistema visual humano.

Como vimos atrás, o tipo de projecção assumida influenciará a forma como se recupera a estrutura. Vamos começar por analisar o caso da projecção ortográfica.

O modelo interno $M(t-1)$ que vai ser utilizado no momento t consiste num conjunto de distâncias entre todos os pares de pontos da estrutura estimada. Formalizando, teremos a seguinte expressão:

$$M(t-1) = d_{ij}(t-1) = D\left[\vec{P}_i(t-1), \vec{P}_j(t-1)\right]; i = 1,\ldots,N; j = 1,\ldots,N; i \neq j \qquad (4.27)$$

Este modelo será imposto na rede através dos pesos $w_{ij}(t)$. Uma vez que cada ponto na estrutura corresponde a um neurónio na rede neuronal, um processo lógico de representar a estrutura seria atribuir aos pesos $w_{ij}(t)$ os valores das distâncias $d_{ij}(t-1)$. No entanto, no caso da projecção ortográfica, pode-se simplificar o processo uma vez que os únicos valores desconhecidos são as coordenadas *Z* (profundidades) dos vários pontos. O custo computacional poderá ser reduzido se, em vez de utilizarmos as diferenças entre todas as coordenadas (*X, Y, Z*), utilizarmos apenas as diferenças entre as coordenadas *Z*. Assim, os pesos poderão ser determinados pela seguinte expressão:

$$w_{ij}(t) = \Delta\vec{Z}_{ij}(t) = \left|\vec{Z}_i(t) - \vec{Z}_{ij}(t)\right| = \sqrt{d_{ij}^2(t-1) - \left[X_i(t) - X_j(t)\right]^2 - \left[Y_i(t) - Y_j(t)\right]^2} \qquad (4.28)$$

O processo de convergência intra-imagem consistirá em computar o valor dos *outputs* $\{Z_i(t); i = 1,\ldots,N\}$ dos vários neurónios que estejam de acordo com a imagem $I(t)$ e com o modelo interno $M(t-1)$. Apesar de, em princípio, a rede estabilizar na

mesma estrutura independentemente das suas condições iniciais, é aconselhável atribuir inicialmente aos *outputs* dos neurónios valores aleatórios para evitar a ocorrência de tendências indesejáveis.

Vamos representar o conjunto dos vizinhos do neurónio i por N_i. Diz-se que dois neurónios são vizinhos se existir uma conexão entre eles. Representemos também por $Z_j, j \in N_i$ o *output* actual do neurónio j. A posição do ponto $\vec{P}_j$ na estrutura será definida em conjunto pelo *output* do neurónio j e pelos valores do seu *input* (X_j, Y_j). Se considerarmos que ao neurónio vizinho j está associado o peso w_{ij}, então a posição do ponto $\vec{P}_i$ deverá ser obtida utilizando uma das seguintes alternativas:

$$Z_i^{j+} = Z_j + w_{ij} \tag{4.29}$$

ou

$$Z_i^{j-} = Z_j - w_{ij} \tag{4.30}$$

Tendo em conta esta informação fornecida pelo neurónio j, existem dois valores igualmente plausíveis para a profundidade do ponto $\vec{P}_i$. De forma semelhante, todos os outros neurónios vizinhos de i contribuem com duas sugestões para a sua profundidade. Partindo destas sugestões, o valor correcto para Z_i deverá ser escolhido, de acordo com o estado actual da rede. Caso todas as posições correntes dos pontos vizinhos $\vec{P}_j$ fossem consistentes com o modelo interno actual, então uma das duas posições sugeridas por cada um dos vizinhos seria igual para todos eles e seria a posição correcta para o ponto $\vec{P}_i$. No entanto, geralmente isso não se passa. O valor de Z_i deverá ser determinado minimizando-se a seguinte expressão:

$$Z_i = \min_Z \left[\sum_{j \in N_i} (Z_i^j - Z)^2 \right] \tag{4.31}$$

onde $Z_i^j = Z_i^{j+}$ ou $Z_i^j = Z_i^{j-}$ e Z representa as várias médias calculadas entre as sugestões dos neurónios vizinhos. Ou seja, escolhe-se a posição média que apresente um menor desvio em relação às sugestões de todos os neurónios vizinhos. É através desta

expressão que também são permitidos alguns desvios da rigidez, por forma a maximizar a consistência entre o modelo interno e a imagem actual.

Resumindo, cada neurónio tenta escolher uma posição para o seu ponto correspondente por forma a satisfazer as restrições impostas pela rede. Portanto, o processo de convergência intra-imagem implica apenas operações locais a nível dos neurónios, enquanto que o esquema de Ullman exigia a optimização de uma função objectivo global. Eventualmente, cada neurónio atingirá um *output* estável, o que conduzirá, por sua vez, à estabilidade da rede e, consequentemente, a uma estrutura tridimensional estável. Esta estrutura está de acordo com a imagem actual mas não é totalmente consistente com o modelo interno corrente. Portanto, será necessário actualizar o modelo interno através de um processo denominado processo de convergência inter-imagens, o qual será descrito de seguida.

O modelo interno $M(t-1)$ é o resultado de todas as computações anteriores ao momento t e reflecte a compreensão que actualmente a rede tem da estrutura tridimensional. Consiste num conjunto de distâncias entre todos os pontos da estrutura e é utilizado para determinar os pesos actuais $w_{ij}(t)$ da rede. A t-ésima imagem será então introduzida como *input* na rede, podendo conduzir a uma nova alteração do modelo interno. A sua introdução irá permitir obter novos *outputs* da rede, os quais serão obtidos através do processo de convergência intra-imagem atrás descrito. Uma vez que $M(t-1)$ não corresponde completamente à estrutura que está a ser observada actualmente, as distâncias entre as coordenadas dos pontos fornecidas pela rede diferirão das distâncias impostas pelo modelo interno. Essa diferença será o mais pequena possível como resultado da aplicação do processo de convergência intra-imagem. Será então necessário actualizar o modelo interno por forma a reflectir o *output* obtido pela rede neuronal. Para esse fim, será utilizada a seguinte expressão:

$$d_{ij}(t) = \sqrt{\left[X_i(t) - X_j(t)\right]^2 + \left[Y_i(t) - Y_j(t)\right]^2 + \left[Z_i(t) - Z_j(t)\right]^2} \qquad (4.32)$$

Por sua vez, este novo modelo interno será utilizado no momento $t + 1$ em conjunto com a nova imagem $I(t + 1)$ para a determinação dos novos pesos $w_{ij}(t+1)$ através da aplicação da expressão (4.28).

Resumindo, o algoritmo deverá seguir as seguintes etapas:

1) Determinar o modelo interno inicial utilizando a primeira imagem da sequência e igualando a zero as diferenças de profundidade (estrutura achatada, sem profundidade). Seria, portanto, utilizada a seguinte expressão:

$$M(1) = \{d_{ij}(t) = \sqrt{\left[X_i(1) - X_j(1)\right]^2 + \left[Y_i(1) - Y_j(1)\right]^2}\,; i, j = 1,\ldots,N; i \neq j\} \quad (4.33)$$

2) Obter a nova imagem $I(t)$ da sequência e introduzir as coordenadas dos pontos característicos nela contidos nos neurónios correspondentes.

3) Tendo M $(t\text{-}1)$ como modelo interno actual, calcular os pesos $w_{ij}(t)$ para cada conexão da rede utilizando a expressão (4.28).

4) Atribuir valores aleatórios pequenos aos *outputs* de todos os neurónios.

5) Seleccionar um neurónio da rede e, para cada um dos seus vizinhos, obter dois valores plausíveis para a profundidade do ponto correspondente utilizando as expressões (4.29) e (4.30). A partir deste conjunto de valores, determinar o novo *output* do neurónio utilizando a expressão (4.31).

6) Repetir o passo 5), efectuando actualizações assíncronas e paralelas dos *outputs* dos neurónios. O processo deverá continuar até que todos os *outputs* estabilizem num valor fixo (processo de convergência intra-imagem).

7) Computar o novo modelo interno $M(t)$ da estrutura tridimensional utilizando a expressão (4.32).

8) Repetir os passos 2) a 7) para todas as imagens da sequência (processo de convergência inter-imagens). *M(t)* representa o modelo interno após *t* imagens terem sido introduzidas na rede. Após uma sequência suficientemente longa, este modelo deverá corresponder à estrutura tridimensional real.

O processo atrás descrito pressupõe a utilização da projecção ortográfica. Se pretendermos utilizar a projecção central, a qual está mais de acordo com imagens reais, será necessário efectuar algumas alterações ao processo descrito.

A projecção dos pontos no plano das imagens será feita de acordo com a expressão (4.26). Para se determinar as coordenadas tridimensionais de um ponto a partir da sua posição na imagem será necessário determinar o parâmetro λ. Assim, o *output* de cada neurónio será constituído pelo valor do λ correspondente.

Outra diferença reside no facto de as diferenças nas profundidades de dois pontos já não poderem ser determinadas directamente a partir das projecções nas imagens, tal como se passava no caso da projecção ortográfica. Assim, os pesos deverão recorrer aos valores das distâncias $d_{ij}(t-1)$ em vez de $\Delta Z_{ij}(t)$. Teremos então a seguinte expressão:

$$w_{ij}(t) = d_{ij}(t-1) \tag{4.34}$$

a qual substituirá a expressão (4.28) no passo 3 do algoritmo. Para além disso, no passo 5, as expressões (4.29) e (4.30) também já não são válidas. As posições plausíveis de um ponto $\vec{P}_i$, sugeridas pelo estado actual dos neurónios vizinhos, por exemplo do neurónio *j*, serão calculadas utilizando a seguinte equação do segundo grau:

$$w_{ij}^2(t) = \left[\lambda_i^j x_i(t) - X_j(t)\right]^2 + \left[\lambda_i^j y_i(t) - Y_j(t)\right]^2 + \left[(\lambda_i^j - 1)f - Z_j(t)\right]^2 \tag{4.35}$$

Uma vez que se trata de uma equação de segundo grau, continuaremos a ter duas posições plausíveis, λ_i^{j+} e λ_i^{j-}, sugeridas por cada um dos neurónios vizinhos.

De resto, o algoritmo manter-se-á idêntico ao caso da projecção ortográfica.

De seguida, iremos efectuar algumas considerações acerca deste algoritmo. Em primeiro lugar, o método descrito é capaz de superar a existência de erros, desde que não muito significativos, aquando do estabelecimento das correspondências. Isso resulta do facto de se utilizar um esquema de rígidez incremental, o qual permite a existência de desvios da rígidez. Quando existam algumas correspondências efectuadas de forma errada em determinado par de imagens, o modelo interno desviar-se-á temporariamente da estrutura correcta mas voltará a convergir quando forem introduzidas novas imagens da mesma parte da cena com correspondências correctas.

O problema da oclusão de pontos característicos, pelos erros que pode gerar, também terá que ser solucionado. Para esse efeito será necessário, antes de mais, detectar quais os pontos que estão oclusos em determinada imagem. Dado o actual modelo interno $M(t\text{-}1)$ e a nova imagem $I(t)$, a detecção de pontos oclusos será efectuada comparando-se os pontos pertencentes ao modelo interno com os pontos que são visíveis na imagem. Todos os pontos pertencentes ao modelo interno que não estejam visíveis na imagem serão pontos sujeitos a oclusão. De seguida, será necessário inibir os neurónios correspondentes a esses pontos para que não possam contribuir para a determinação do *output*. As distâncias que envolvam pontos oclusos permanecerão inalteradas no modelo actual. Quandos os pontos voltarem a ser visíveis, os neurónios serão novamente activados.

Laganière e Cohen efectuaram várias experiências para testarem o seu algoritmo, tendo chegado a algumas conclusões.

Em primeiro lugar, um aumento no número de pontos da estrutura atrasa um pouco a convergência para a estrutura tridimensional, não afectando significativamente, no entanto, o processo de convergência intra-imagem. Foi notado que, normalmente, a rede estabiliza depois de cada neurónio ter sido actualizado cinco ou seis vezes. No entanto, a complexidade da actualização de cada neurónio depende do número dos seus vizinhos, ou seja, do seu número de conexões. Não é necessário que cada neurónio esteja conectado a todos os restantes neurónios da rede. Laganière e Cohen notaram que o facto de se utilizar um número menor de neurónios vizinhos não afecta a convergência do algoritmo. No entanto, será necessário estabelecer conexões por forma a manter uma estrutura globalmente conexa.

Comparando o seu algoritmo com o esquema de Ullman, estes investigadores chegaram à conclusão que, para estruturas com poucos pontos, o esquema de Ullman é mais rápido a convergir porque existem poucos pontos a considerar na minimização da função objectivo. No entanto, quando o número de pontos aumenta, a rede neuronal é claramente superior pois tira partido do seu processamento em paralelo. Isto faz com que este último método seja mais indicado para lidar com imagens reais, pois estas normalmente apresentam estruturas complexas.

5. Conclusões

Uma primeira conclusão que se pode retirar da análise por nós efectuada, no que diz respeito às redes neuronais, é que estas apresentam certas características que as tornam muito atractivas para a resolução de certos problemas. Nomeadamente, permitem um processamento de informação em paralelo. Todos os neurónios trabalham simultaneamente para a obtenção dos mesmos fins. No entanto, caso um deles falhe, esse facto não irá afectar o desempenho da rede, desde que os restantes continuem a funcionar. Por outro lado, ao contrário dos processos sequênciais, não é necessário especificar regras e procedimentos para a resolução de problemas. Tudo o que é necessário efectuar é treinar a rede com alguns exemplos para que ela possa aprender a dar as respostas correctas. Esta sua capacidade de aprendizagem é muito importante, uma vez que lhes permitirá efectuarem generalizações para novos casos que nunca tenham encontrado anteriormente.

Quanto à utilização do modelo de rede neuronal de Hopfield para o estabelecimento de correspondências, este apresenta a vantagem, em relação aos modelos tradicionais, de podermos introduzir todas as restrições do problema numa função de energia, o que se torna muito cómodo. A rede encarregar-se-á de impôr essas restrições através da minimização da função de energia. O facto de esta rede apresentar uma arquitectura com todos os neurónios interconectados entre sí, cada um recebendo e fornecendo informação a todos os restantes, tornam a computação rápida e potente. Teremos assim um estabelecimento de correspondências efectuado de forma automática. Em contrapartida, os investigadores de redes neuronais já notaram que as redes de Hopfield, nalguns casos, não funcionam muito bem como instrumento de optimização, principalmente nos casos em que o número de variáveis seja muito elevado.

Por outro lado, o modelo de rede neuronal utilizado para a obtenção da estrutura tridimensional apresenta a vantagem de tolerar, até certo nível, a existência de erros na fase anterior de estabelecimento de correspondências. Esses erros provocarão um desvio temporário da estrutura em relação à estrutura tridimensional correcta mas aquela irá convergir novamente de forma correcta quando voltarem a ser introduzidas novas correspondências bem estabelecidas. Para além disso, o processo de convergência para a estrutura tridimensional correcta não é significativamente afectado pelo aumento do

número de pontos na estrutura a analisar. Esta característica é particularmente útil para a análise de imagens reais, uma vez que, normalmente, as estruturas nelas presentes são compostas por muito pontos. Nos casos em que o número de pontos é elevado, a rede neuronal também apresenta um desempenho muito superior ao Esquema da Rígidez Incremental de Ullman, uma vez que tira partido da sua capacidade de processamento em paralelo. As operações com vista à optimização são levadas a cabo a nível local nos neurónios, enquanto que no algoritmo de Ullman era necessário efectuar uma optimização a nível global, tornando o processo mais complexo. Para além disso, a complexidade computacional da rede pode ser diminuída através da utilização de menos conexões entre os neurónio. Desde que a sua estrutura se mantenha conexa, a rede continuará a convergir para a solução correcta.

Somos de opinião que o esquema baseado em redes neuronais por nós proposto é suficientemente robusto para solucionar o problema abordado nesta dissertação. Essa robustez resulta principalmente das vantagens decorrentes da utilização de redes neuronais, as quais já foram referidas atrás. Assim sendo, pensamos que atingimos o objectivo que nos propusemos alcançar, ou seja, definir um modelo que permitisse recuperar a estrutura tridimensional de uma cena a partir da análise de uma sequência de imagens contendo projecções bidimensionais dessa mesma cena. Este modelo poderá ser utilizado em aplicações no domínio da Realidade Virtual nas quais seja necessário obter informações acerca da tridimensionalidade de determinado ambiente, como é o caso de sistemas de Telepresença.

O método por nós proposto não pode ser aplicado em tempo real. Tal facto resulta da grande complexidade computacional deste problema. O estado actual da tecnologia não permite uma computação suficientemente rápida. No entanto, o constante desenvolvimento verificado na potência de processamento dos computadores leva a crer que um dia poderemos dispôr de máquinas suficentemente rápidas. Por outro lado, o processo sequêncial que os actuais computadores seguem na computação da informação torna-os pouco apropriados para a sua utilização na implementação de modelos baseados em redes neuronais. Esta situação poderá modificar-se com o advento dos computadores com processamento em paralelo e com a tecnologia VLSI. Somos de opinião de que nessa altura o modelo por nós proposto poderá ser facilmente implementado em tempo real, possibilitando grandes ganhos ao nível da interacção do utilizador com o meio ambiente analisado.

Outra conclusão a que chegamos é que a RV apresenta todas as características necessárias para ser considerada como uma TI. De facto, a RV, através da utilização de um conjunto de tecnologias, é capaz de criar, armazenar e distribuir informação. Para além disso, aínda apresenta uma série de vantagens em relação às TI tradicionais. Essas vantagens traduzem-se na obtenção de informação mais adequada e de uma forma mais eficaz. Os sistemas de RV podem ser concebidos por forma a que somente a informação relevante seja apresentada ao utilizador. Nos casos em que exista imersão, o utilizador fica isolado de factores externos que potencialmente possam afectar a sua concentração. Por outro lado, o facto de a informação ser convertida em metáforas visuais permitirá que a sua assimilação seja feita de uma forma mais intuitiva. O cérebro tem uma maior facilidade no tratamento de informação visual do que informação na forma de texto ou números. O reconhecimento de formas visuais é feita de uma forma praticamente automática pelo cérebro, permitindo uma maior rapidez de assimilação da informação por parte do utilizador.

Por outro lado, a RV vem possibilitar aos seus utilizadores uma interacção sem precedentes com a informação. Através de equipamento adequado, será possível interagir com a representação tridimensional da informação. Este processo permite visualizar imediatamente quais os efeitos das acções desenvolvidas pelo utilizador, permitindo-lhe uma interacção em tempo real com a informação. Esta característica torna a RV óptima para todo o tipo de simulações. Para além disso, o utilizador poderá escolher, de entre as várias formas possíveis de visualizar a informação, aquela que se adapte melhor às suas necessidades. Isto é muito importante, uma vez que permite adaptar a informação ao utilizador em vez de ser este a adaptar-se à informação, algo que só é possível devido à grande interactividade que a RV possibilita.

Ao nível das UE, tal como na sociedade em geral, verificar-se-ão efeitos significativos provocados pela introdução da RV. A informação será tratada de uma forma completamente nova. A sua utilização conduzirá à obtenção de informação mais eficiente e de uma forma mais adequada às suas necessidades, permitindo a obtenção de grandes ganhos que poderão conduzir à existência de vantagens competitivas sobre outras UE que não utilizem a RV. As possibilidades de aplicação são enormes, desde o topo até ao mais baixo nível da estrutura organizacional. De uma forma geral, permitirá o desempenho de várias tarefas organizacionais de uma forma mais eficiente e eficaz.

Por outro lado, poderá modificar totalmente a forma como todos nós trabalhamos. A Telepresença poderá eliminar, pelo menos em parte, a necessidade de reunir trabalhadores no mesmo espaço físico. Será possível trabalharmos a partir das nossas casas, ligados por rede às instalações virtuais do local de trabalho. A interação com o ambiente de trabalho e com as pessoas nele presentes seria efectuada à distância. Há uma enorme quantidade de possibilidades que se abrem, as quais terão que ser exploradas sob o risco de perda de competitividade.

De facto, a RV possui elevados potenciais. No entanto, esses potenciais ainda não podem ser totalmente explorados. Tal facto deve-se ao estado actual das tecnologias por ela utilizadas, que é insuficiente para apresentar o desempenho que dela se espera. Por um lado, a nível gráfico, os computadores não são suficientemente rápidos para debitarem imagens realistas e actualizarem de forma rápida os écrans. Os próprios monitores utilizados nos capacetes apresentam resoluções muito baixas e ângulos de visão reduzidos. É ao nível do tacto que os dispositivos utilizados se encontram menos desenvolvidos. De facto, é extremanente difícil simular sensações tácteis. Para além disto, os dispositivos utilizados para efectuar a interacção ainda não são muito práticos e, em termos gerais, o equipamento utilizado ainda não se encontra miniaturizado, constituíndo um certo incómodo para o utilizador. Por exemplo, os capacetes são demasiado grandes e pesados, tornando a sua utilização desconfortável.

Por outro lado, esta tecnologia ainda não atingiu o ponto de estar disponível para a sociedade em geral. Os seus preços são proibitivos. A inexistência de *standards* definidos não permite ainda uma uniformização e uma consequente produção em massa. Existem muitas indefinições que será necessário ultrapassar para que seja possível adoptar padrões universais e começar-se a produzir em série. Quando tal ocorrer, verificar-se-á uma queda nos preços, tornando a RV acessível à generalidade dos utilizadores.

Como conclusão final, pensamos que as tecnologias de informação por nós analisadas (Redes Neuronais e Realidade Virtual) apresentam características muito próprias e de grande impacto futuro que, quando o desenvolvimento tecnológico permitir a sua utilização em massa, irão alterar radicalmente o actual panorama das Tecnologias de Informação e dos Sistemas de Informação nas Unidades Económicas.

Referências Bibliográficas

Burns, J.; Hanson, A.; Riseman, E.; (1986); "Extracting Straight Lines"; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAM-8, No. 4, pag. 425-455, July 1986

Dhond, U.; Aggarwal, J.; (1995); "Stereo Matching in the Presence of Narrow Occluding Objects Using Dynamic Disparity Search"; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 7, pag. 719-724, July 1995

Erten, G.; Goodman, R.; (1996); "Analog VLSI Implementation for Stereo Correspondence Between 2-D Images"; *IEEE Transactions on Neural Networks*, Vol. 7, No. 2, pag. 266-277, March 1996

Grimson, W.; (1981); "A Computer Implementation of a Theory of Human Stereo Vision"; *Phil. Trans. Royal Soc. London*, Vol. B292, pag. 217-253

Grimson, W.; (1985); "Computational Experiments With a Feature-Based Stereo Algorithm"; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, No. 1, pag. 17-34, January 1985

Haykin, S.; (1994); "Neural Networks: A Comprehensive Foundation"; Macmillan College Publishing Company; New York

Hertz, J.; Krogh, A.; Palmer, R.; (1991); "Introduction to the Theory of Neural Computation"; Addison-Wesley Publishing Company, Massachusetts

Hopfield, J.; (1982); "Neural Networks and Physical Systems with Emergent Collective Computational Abilities"; *Proceedings of the National Academy of Sciences*, USA Vol. 79, pag. 2554-2558

Hopfield, J.; Tank, D.; (1985) "Neural Computation of Decisions in Optimization Problems"; *Biological Cybernetics*, Vol 52, pag. 141-152

Hsu, J.; Kusnan, J.; (1989); "The Fifth Generation: The Future of Computer Technology"; Windcrest, Blue Ridge Summit

Laganiere, R.; Cohen, P.; (1995): "Gradual Perception of Structure from Motion: A Neural Approach"; *IEEE Transactions on Neural Networks*, Vol. 6, No. 3, pag. 736-748, May 1995

Lew, M.; Huang, T.; Wong, K.; (1994); "Learning and Feature Selection in Stereo Matching"; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 9, pag. 869-881, September 1994

Marr, D.; (1982); "Vision"; W.H. Freeman and Co., San Francisco

Moravec, H.; (1981); "Robot Rover Visual Navigation"; Ann Arbor, MI: U.M.I.; Research Press

Nasrabadi, N.; Choo, Y.; (1992); "Hopfield Network for Stereo Vision Correspondence"; *IEEE Transactions on Neural Networks*, Vol. 3, No. 1, pag. 5-13, January 1992

Oliveira, E.; Ramos, C.; (1988); "Uso de Modelos Flexíveis e Simbólicos na análise de uma Cena"; *Actas do 5.º Congresso Português de Informática*, Comunicação N.º 21, pag. 356-369; Associação Portuguesa de Informática

Pimentel, K.; Teixeira, K.; (1993); "Virtual Reality: Through the New Looking Glass"; Windcrest Books

Poggio, T.; (1984); "Vision by Man and Machine"; *Scientific American*, Vol. 250, pag. 106-115, April 1984

Rumelhart, D.; Hinton, G.; Williams, R.; (1986); "Learning Representations by Back-Propagating Errors"; *Nature*, No. 323, pag. 533-536

Senn, J.; (1996); "Information Technology in Business: Principles, Practices, and Opportunities"; Prentice-Hall International, Inc.

Sharda, R.; (1994); "Neural Networks for the MS/OR Analyst: An Application Bibliography"; *Interfaces*, Vol. 24:2 March-April 1994, pag. 116-130

Ullman, S.; (1984); "Maximizing Rigidity: the Incremental Recovery of 3-D Structure From Rigid and Nonrigid Motion"; *Perception*, Vol. 13, pag. 255-274

Weng, J.; Ahuja, N.; Huang, T.; (1992); "Matching Two Perspective Views"; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 8, pag. 806-825, August 1992

Weng, J.; Huang, T.; Ahuja, N.; (1987); "3-D Motion Estimation, Understanding and Prediction from Noisy Image Sequences" ; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 3, pag. 370-389, May 1987

Referências Bibliográficas Históricas

Marr, D.; Poggio, T.; (1976); "Cooperative computation of Stereo Disparity"; *Science*, Vol. 194, pag. 283-287

Marr, D.; Poggio, T.; (1979); "A Computational Theory of Human Stereo Vision"; *Proc. Royal Soc. London*, Vol B204, pag. 301-328

McCulloch, S.; Pitts, W.; (1943); "A Logical Calculus of Ideas Immanent in Nervous Activity"; *Bulletin of Mathematical Biophysics*, Vol. 5, pag. 115-133

Moravec, H.; (1977); "Towards automatic visual obstacle avoidance"; *Proc. 5th Int. Joint Conf. Artificial Intell.*; p. 584

Rosenblatt, F.; (1958) "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain"; *Psychological Review*, No. 65, pag. 386-408

Widrow, B.; Hoff, M.; (1960); "Adaptative Switching Circuits"; In *1960 IRE WESCON Convention Record*, part 4, pag. 96-104, New York, IRE